EXECUTIVE FUNCTION MEASURES IN EARLY CHILDHOOD SCREENING: CONCURRENT AND PREDICTIVE VALIDITY

A Dissertation
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Amanda Wenzel Kalstabakken

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILSOPHY

Advisor: Ann S. Masten, Ph.D.

June 2017

Acknowledgment

This research was supported with graduate fellowships from the National Science Foundation and the University of Minnesota as well as by a research grant to Ann Masten from the Fesler-Lampert Chair in Urban and Regional Affairs. This project represents a collaborative effort with the University of Minnesota and Minneapolis Public Schools. I am incredibly grateful for the children and families who participated in this project and for the efforts of the following individuals who contributed:

Minneapolis Public School Administrative Staff: Cynthia Hillyer, Maureen Seiwart, Ryan Strack, Melody Jacobs-Cassuto, Doug Marston, Amanuel Medhanie, Elizabeth Hinz

Minneapolis Public Schools Screening Staff: Peter O'Gormon, Zenebech Demissie, Mariam Warsame, Ker Vue, Charletta Mosely, Maria Doerr, Marilyn Selb, Zuleika Severini, Iftu Hunte, Marvin Lyles, Kathy Cromie, Julie Streitz, Delphie Sorenson, Diane Bonniwell, Patricia Dowdle, Diane Hite, Pat Skirka

University of Minnesota Collaborators: Ann Masten, Christopher Desjardins, Stephanie Carlson, Philip Zelazo

University of Minnesota Graduate Students and Research Assistants: Jacob Anderson, Katie Berghuis, Sandy Moran, Maya Buckner, Mai Tong Yang, Mariah Geiger, Ambrosia Smith

In addition to those who contributed directly to this project, I am deeply grateful for those who have supported me throughout my graduate career. First and foremost, I am thankful for the opportunity to have studied with Ann Masten. Her ability to teach and lead has consistently been accompanied by unfaltering compassion, a sense of humor, and a strong sense of joy in pursuing meaningful research. I am further thankful for the opportunity work alongside my cohort and fellow lab mates. In addition to those already mentioned, special thanks to J.J. Cutuli, Aria Fiat, Megan Finsaas, Catrina Helseth, Janette Herbers, Zoe Jacobson, Madelyn Labella, Theresa LaFavor, Katie Lingras, Sandra Ahumada Lopez, Christopher McCormick, Jill Merrick, Amy Monn, Angela Narayan, Laura Supkoff Nerenberg, Julianna Sapienza, Meg Soli, and Eric Thibodeau.

In addition to my research training, I have been fortunate to work with my clinical supervisors, who not only helped me develop as a clinician but always had open doors as I navigated through various milestones of my graduate program. I would like to especially thank Jeffery Wozniak, Sasha Zagoloff, Timothy Piehler, Amy Gross, and Christopher Boys.

Lastly, I would like to thank my friends and family. To my parents, Barb and Bill Wenzel, and my brother, Ryan Wenzel, thank you for cheering me on through each step and never tiring. To my husband, Kyle, I am lucky to have had you through this entire journey. Thank you for always being beside me and reminding me to celebrate and enjoy even the small moments. I cannot imagine this process without your patience, love, and support.

To Kyle.

Abstract

Recently, evidence has emerged suggesting that executive function (EF) in early childhood is related to multiple aspects of school success, including learning, academic achievement, and social functioning. Despite this evidence, little or no research has focused on the value of EF assessment in routine early childhood assessments. The present study, with the collaboration of a large and diverse urban school district, examined the concurrent and predictive validity of EF assessment in the context of screening. Analyses focused on a sample of 461 children ages 3 to 5 years old who were assessed on three EF tasks in addition to routine screening. EF measures included two computerized NIH Toolbox measures, the Flanker and Dimensional Change Card Sort (DCCS) tasks, with developmental extensions (Dext) that lower the floor of these measures, and Peg Tapping. Routine screening included well-validated developmental readiness and behavioral measures. Results of multivariate analyses indicated that Flanker-Dext and DCCS-Dext worked well to lower the floor of the Toolbox EF measures and that the three EF tasks showed the expected convergent validity. Structural equation models indicated that a latent variable based on the three EF scores was highly related to performance on the developmental screener and moderately and negatively related to other screening measures, including parent-report of child social-emotional problems and screener-observed behavior problems. EF was a significant predictor of academic skills at the beginning of kindergarten and predicted growth across the school year. EF did not provide unique predictive value for predicting academic outcomes over and above other screening measures in use by this district. In contrast, EF both contributed unique predictive value for kindergarten behavior and also was the only measure administered during screening that continued to have unique predictive value for first-grade behavior (over kindergarten behavior). These findings indicate that EF is important for school success, can be assessed by brief and easy-to-learn measures during screening, and may be an important target for intervention in early childhood.

Table of Contents

List of Tables

List of Figures

Introduction

For the past several decades, there has been an increased focus on educational policy surrounding school readiness to ensure that children are ready to learn at the beginning of kindergarten (Kagan & Kauerz 2007; Kagan & Rigby, 2003; National Education Goals Panel, 1991). Although most agree that a readiness to learn is important for school success, it is unclear what, specifically, this entails or how to best measure it (Boan, Aydlett, Multunas, 2007; Kagan, 1990). This is in part because the kindergarten year represents a significant, and for some children difficult, transition into a formal learning environment that has both increased learning and behavioral expectations (Rimm-Kaufman & Pianta, 2000). As such, school readiness can be viewed as a multi-faceted construct (Blair, 2002; Eisenberg, Valiente, & Eggum, 2010)

This burgeoning interest in school readiness has led to an increased effort in assessing young children prior to the start of school to identify those who are atypically developing and therefore at-risk for school failure (reviewed in Boan et al., 2007). The practice of assessing children in early childhood became more widespread in 1975 when the Individuals with Disabilities Education Act (Public Law 94-142) was passed by the U.S. Congress, which allowed access to public services for early identification of children with developmental disabilities. In 1986, Public Law 99-457 extended these services to children between 3 and 5 years old. With the passing of these laws, young children from diverse backgrounds, including those from low-income families, were afforded the opportunity to be assessed and receive services.

Often, the process of identifying children who need further services involves developmental screening, in which children are briefly assessed with respect to general knowledge, language, physical health, and occasionally social-emotional functioning (Boan et al., 2007; Brassard & Boehm, 2007). Early identification of developmental concerns, such as during the preschool years, provides an opportunity to remediate problems and address the needs of children and families prior to the start of school. Access to services during this time may be especially critical as the preschool years represent a period of rapid cognitive development (Carlson et al., 2013; Kochanska, Coy, Murray, 2001; Zelazo & Muller, 2002), allowing for a high return on invested services. This is supported by evidence that early intervention programs can have a positive impact on children's cognitive and social-emotional development (e.g., Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001; Ramey, Pungello, Sparling, & Miller-Johnson, 2002; Ramey et al., 2000).

Despite increased efforts to identify children in need of services, many children still have difficulty with the transition to school (Rimm-Kaufman, Pianta, & Cox, 2000), which has prompted beliefs that readiness for school may extend beyond general knowledge and social-emotional development to include aspects of child functioning like self-regulation and executive functioning (EF) (Blair, 2002; Eisenberg et al., 2010). Although there is strong evidence supporting that EF is important for later academic achievement (e.g., Bull & Scerif, 2001; Bull, Espy, & Wiebe, 2008; Clark, Pritchard, & Woodward, 2010; Masten et al., 2012; Mazzocco & Kover, 2007) and classroom behavior (e.g., Bierman, Nix, Greenberg, Blair & Domitrovich., 2008; Masten et al., 2012; Obradovic, 2010; Wilson, 2003), there are no studies to the author's knowledge that incorporate EF assessment into early childhood screening.

The following literature review provides a brief overview of common assessment approaches used prior to or at school entry and describes the existing literature supporting the inclusion of EF measures. This dissertation, which is part of a larger, ongoing study on the potential of EF measures for early childhood screening, had several overarching goals. First, this study was designed to evaluate two measures of EF from National Institutes of Health (NIH) Toolbox that were modified to include developmental extensions. These developmental extensions, which were embedded in the regular NIH Toolbox Measures, are intended to lower the floor of the tasks so that children from more diverse backgrounds or with lower EF skills can attain valid scores. When individuals perform poorly during practice or the initial trials of either task, the developmental extension module is triggered automatically (See Appendices A and B for more detail). The present study examined the frequency with which either extension was needed in the context of a routine screening in an urban school district, where the screening population has high proportions of disadvantaged children at risk school achievement or adjustment problems.

Second, this dissertation study examined the convergent validity of the EF measures, both with respect to their covariance with each other and also in regard to their relations with other variables that would be expected to covary with EF, such as age and socioeconomic status (SES).

Third, the study examined the concurrent validity of EF as an indicator of developmental readiness by observing how well EF scores relate to well-validated screening measures already in use by the collaborating school district as well as this region more broadly.

Finally, the predictive value of EF scores assessed in screening with respect to both kindergarten academic achievement and early school behavior problems was examined by assessing how EF at screening relates to school outcomes.

Identifying Children At-Risk for School Failure

The two commonly used strategies for identifying children's readiness to learn are developmental screening and school readiness assessments. Developmental screenings are intended to be brief, such that they can be administered on a large-scale basis, and are used to identify areas of children's development that need further evaluation or diagnostic assessment (Boan et al., 2007; Brassard & Boehm, 2007). Children are typically divided into groups, such as (a) refer for more thorough evaluation, (b) rescreen, or (c) no problems/pass (Brassard & Boehm, 2007). More thorough evaluation can help determine what interventions, if any, are most appropriate. This includes possibly referring children for special education services (Meisels, 1999).

School readiness assessments, on the other hand, aim to identify children's school-related skills and broadly represent a child's preparedness to participate in a given curriculum (Boan et al., 2007; Brassard & Boehm, 2007; Kagan & Rigby, 2003; Meisels, 1999). Ultimately, these measures aim to identify areas of strength or weakness for the child. Outcomes from these assessments may inform school approaches to curricula development as well as highlight any individualized programming needs that children have (Mehaffie & McCall, 2002).

Despite the fact that these two types of assessments are intended to be used for different purposes, there is often significant overlap between the content of the measures. For example, developmental screenings often include domains such as physical, cognitive, language, social-emotional, and adaptive development (Boan et al., 2007; Brassard & Boehm, 2007; Rydz, Shevell, Majnemer, & Oskoui, 2005), while readiness assessments focus on measuring physical development, social-emotional development, general knowledge, approach to learning, and language development (Boan et al., 2007; Kagan, 1992; Kagan & Rigby, 2003; National Education Goals Panel, 1995).

There is strong evidence indicating that areas commonly assessed through early childhood assessments are indeed predictive of later academic outcomes (see Brassard & Boehm, 2007 for review). For example, aspects of children's early language such as phonological sensitivity, alphabet knowledge, vocabulary, and syntax are related to children's later reading comprehension (Lonigan, Burgess, & Anthony, 2000; National Early Literacy Panel, 2005). Further, children's mathematics skills, such as knowledge of numbers and ordinality, at school entry were both predictive of later math and reading achievement (Duncan et al., 2007).

It is important to stress, however, that for both types of assessment the results of testing should only be used for the purposes for which they have been validated. Historically, these measures have been used inappropriately and have been the primary source informing high-stakes decisions for children, such as delaying the start of kindergarten and placement into a readiness kindergarten classroom (Gredler, 1997; May & Kundert, 1992; Meisels, 1987).

Executive Function and School Adjustment

Despite efforts to identify children who are at-risk for school difficulties prior to beginning kindergarten, many children still struggle with the transition to formal schooling. Rimm-Kaufmann et al. (2000) surveyed teachers who reported that approximately half of children demonstrate problems of mild to significant concern with the transition to kindergarten. One of the two areas that teachers endorsed as most problematic was difficulty following directions. Reports such as these suggest that school readiness expands beyond academic skills and knowledge, which have led to increased interest in constructs related to self-regulation, particularly EF.

EF consists of cognitive and behavioral skills involved in directing thoughts and organizing behavior in a goal-oriented manner and includes inhibitory control (i.e., inhibiting a prepotent response in favor of an alternative response), cognitive flexibility, selective attention, and working memory (Barkley, 2001; Carlson, Zelazo, & Faja, 2013). Although EF continues developing across childhood, it develops most rapidly during the preschool years (Carlson et al., 2013; Diamond, 2002; Zelazo & Muller, 2002). In the classroom, EF manifests as children's ability to pay attention, stay on task with completing activities and homework assignments, and comply with instructions. Recent evidence suggests that children with better EF skills may be able to better attend and subsequently retain more information compared to their lower EF peers (e.g., Benson et al., 2013; Hassinger-Das, Jordan, Glutting, Irwin, & Dyson, 2014; Zaitchik, Iqbal, & Carey, 2014). Children with stronger EF skills may also view themselves as more capable learners, which in turn further promotes motivation and engagement in the classroom (Blair, 2002). Self-regulatory skills may also promote learning through their positive influences on peer and teacher relationships (Eisenberg et al., 2010).

Indeed, initial evidence indicates that EF is foundational to later school success. EF skills are predictive of school-related outcomes, such as school readiness (Espy et al., 2004; McClelland et al., 2007; Morrison, Ponitz, & McClelland, 2010; St Clair-Thompson & Gathercole, 2006) and academic achievement (e.g., Bull & Scerif, 2001; Bull et al., 2008; Clark et al., 2010; Masten et al., 2012; Mazzocco & Kover, 2007). Specifically, Bull and colleagues (2008), found that preschool children with better EF skills were at an immediate advantage in reading and math upon starting kindergarten compared to children with poorer EF, with this advantage persisting until at least third grade. Further, results from a randomized-control intervention for children receiving a school readiness program within their Head-Start classroom demonstrated that children with better EF skills at the beginning of the program made greater gains in language and emergent literacy skills than their peers (Bierman et al., 2008).

There are also longer-term impacts associated with early difficulties with learning-related skills, such executive functioning. McClelland, Acock, and Morrison (2006) demonstrated that deficits with these skills were both associated with poorer academic performance in reading and mathematics in kindergarten and that growth curves indicated that this gap widened until second grade and then persisted from third through sixth grade. Individual differences in EF during childhood have also been found to relate to later school success, such as completing high school (Vitaro, brendgen, Larose, & Tremblay, 2005) and college (McClelland, Acock, Piccinin, Rhea, & Stallings, 2013).

In addition to EF skills predicting academic success, children with better EF have more success with aspects of social functioning. Children with stronger EF skills have been found to be more capable of inhibiting reactive aggressive tendencies in favor of more appropriate behavioral responses (Cole, Usher, & Cargo, 1993). Additionally, they show fewer behavioral problems and better social emotional competence (Bierman et al., 2008; Masten et al., 2012; Obradovic, 2010), are more likely to be nominated as well-liked by their classmates (Wilson, 2003), and are better able to cope after experiencing a social failure (Wilson, 2003). This pattern has been found in non-Western cultures as well, with children's poor EF skills – as reported by teachers or parents – having more negative peer evaluations (Eisenberg, Pidad, & Liew, 2001).

Executive Function as a Target for Intervention

Part of the interest in assessing EF during early childhood screening or readiness assessments is due to the evidence that EF is malleable and that preschool represents a time of rapid cognitive development as well as a period of plasticity (see Carlson et al., 2013; Diamond & Lee, 2011; Diamond & Ling, 2016). Studies have demonstrated that EF skills can be improved through individualized training (Bergman Nutley et al., 2011; Espinet, Anderson, & Zelazo, 2013; Holmes, Gathercole, & Dunning, 2009; Holmes et al., 2010; Klingberg et al., 2005; Klingberg, 2010; Kendler & Kendler, 1961; Mischel & Patterson, 1976; Thorell, Lindqvist, Bergman Nutley, Bohlin, & Klingberg, 2009) as well as through classroom-based interventions (Bodrova & Leong, 2007; Bierman et al., 2008; Raver et al., 2009).

In addition to being a malleable target, there is evidence that EF skills are particularly important for children from high-risk backgrounds including homeless/highly mobile (HHM), poverty, early institutional care, and maltreatment (Hackman & Farah, 2009; Loman, Johnson, Westerlund, Nelson, & Gunnar, 2012; Masten et al. 2012). For example, EF assessed prior to the school year was a unique predictor of academic success, beyond general intellectual function and parenting quality, among kindergarten and first-grade children who were experiencing homelessness (Herbers, Cutuli, Supkoff, Narayan, & Masten, 2014; Masten et al., 2012; Obradovic, 2010).

Despite these skills having a unique protective effect, it is important to note that the development of EF is vulnerable to chronic stress (Pechtel & Pizzagalli, 2011). Thus, children who have experienced poverty and adversity in early childhood often show deficits in these skills (e.g., Blair & Raver, 2015; Blair & Razza, 2007; Buckner, Mezzacappa, & Beardslee., 2003; DePrince, Weinzierl, Combs, 2009; Loman et al., 2012; Pears, Fishers, Bruce, Kim, Yoerger, 2010). Given that young children from high-risk backgrounds show a “readiness gap” at the start of kindergarten (Princiotta, Flanagan, & Germino-Hausken, 2006), targeting and attempting to boost known protective factors for these children is critical. In addition, there is evidence that achievement gaps between HHM students and lower-risk groups (e.g., children receiving free meals, reduced meals, or no assistance) tend to persist and even widen across the school years (Cutuli et al., 2013; Herbers et al., 2012), making early identification and intervention even more pressing.

Psychometric Properties of Screening Measures

Because early childhood assessments are widely used to inform important educational decisions, the National Research Council (2008) emphasized that it is critical to choose quality measures when assessing young children. The literature reviewed above suggests a strong link between EF and later school achievement. Below are specific considerations for assessing whether measures demonstrate sufficiently strong psychometric properties to be included in developmental screening.

A technical aspect of a measure that must be considered in determining the test adequacy involves test floors. The test must have a sufficient range of scores in order to differentiate between children (reviewed in Boan et al. 2007). Due to both EF developing rapidly in the preschool years (Carlson et al., 2013; Diamond, 2002; Zelazo & Muller, 2002) and children from high-risk environments being vulnerable to impairments in EF (DePrince et al., 2009; Loman et al., 2012; Pears et al., 2010; Pechtel & Pizzagalli, 2011), it is particularly important that any EF measure used in the context of screening be developmentally sensitive. The current study, in part, addressed this need – highlighted by Carlson (2005) – for more developmentally sensitive measures of EF by examining the range of skills captured by two newly developed measures with a racially, ethnically, and socio-economically diverse sample.

In addition to being developmentally sensitive, measures used during early childhood assessment should have strong psychometric properties. As summarized in Brassard & Boehm (2007), the necessary strength of the psychometric properties will vary as a function of the purpose of the assessment, with higher stakes necessitating stronger psychometric properties. Measures should be evaluated in regards to test-retest reliability, construct validity, and predictive validity. Since the current study focuses on aspects of construct validity and the predictive validity of executive functioning measures administered during early childhood screening, this section review will focus on these psychometric properties in particular.

Construct validity refers to the extent to which an assessment is measuring the construct of interest. One method for gathering evidence about construct validity is through identification of convergent validity (i.e., examining the extent to which one measure operationalizes in a manner that is similar to a theoretically related construct). Brassard and Boehm (2007) recommend using measures that have a correlation coefficient of about .60 with other measures intended to assess the same construct. Additionally, there should be established predictive validity that demonstrates the extent to which a measure predicts a future area of interest, such as whether a child is identified as having a learning difficulty or how a child performs on end of the year mathematics or reading tests (Brassard & Boehm, 2007). Although established predictive validity is important, ideally with a correlation coefficient of .80 with outcome measures of interest, the authors acknowledge that it is unlikely for even a strong assessment to have near perfect predictive validity in regards to children's end of year performance. Beyond the skills commonly assessed in screenings, children's motivation, parental involvement, teacher's skill level, children's attendance, and frequency of school changes will all impact learning across the school year (Brassard & Boehm, 2007).

Present Study

The current study is part of a collaborative, longitudinal project with the Minneapolis Public School (MPS) district. Minnesota State Law (MN Statute 121A.17) requires that children be screened at least once prior to the start of kindergarten, ideally targeting children ages 3 and 4 years. In order to address this need, MPS offers free early childhood screening for 3-to-5-year-old children to identify concerns related to development and school readiness. Children who have received a comparable developmental screening through their primary care provider or other health care provider may be exempt, while children who do not present for screening prior to the start of kindergarten are assessed at the beginning of their kindergarten school year. The current study represents the first known effort to include EF measurement in early childhood screening for children 3 to 5 years of age.

The MPS district utilizes a developmental screener, which is widely used in this region (Minneapolis Preschool Readiness Instrument, Revised; MPSI-R Technical Manual, 2007) and aims identify children who are atypically developing and therefore at-risk for later school problems. Based on the intended use of this measure, this study will refer to the construct measured by the screener as “developmental readiness,” in order to capture that the focus is measuring key aspects of child development that relate to early school functioning.

This dissertation study has four primary aims. The first two aims involve the newly modified EF measures from the NIH Toolbox (see Appendices A and B) the Dimensional Change Card Sort with Developmental Extension (DCCS-Dext; Carlson, Zelazo, Anderson, Kalstabakken, & Masten, 2015) and the Flanker – Developmental Extension

(Flanker-Dext; Anderson, Zelazo, Carlson, Kalstabakken, & Masten, 2015). As noted above, these extensions which were created by a team at the University of Minnesota to address the need for developmentally sensitive EF measures, extend the floor of the measures. The first aim was to assess whether these measures capture a greater range of performance in young children,

It was hypothesized that:

a. The NIH Toolbox EF measures with developmental extensions would be able to capture a wider range of skill levels, with no observable floor or ceiling effects. It was expected that many children would require the developmental extension, especially the children from higher-risk backgrounds (i.e., HHM).

The second aim of this study was to assess the convergent validity of the two computerized EF measures by examining the covariance of the two tasks with each other and also with another widely-used measure of EF for the same age level (Peg Tapping). In addition, the correlation of the EF scores with other well-established demographic variables, such as age and SES, were tested, also to assess convergent validity.

Specific hypotheses corresponding with the second aim were as follows:

b. The two newly developed measures of EF would be positively related to demographic factors that are known to be related to EF.

c. The three EF tasks would be moderately and positively related to one-another, even after controlling for age.

The third aim involved assessing the construct validity of EF as an indicator of developmental readiness. The following hypotheses were tested:

a. EF performance would be moderately and positively related to performance on the developmental screener beyond what is expected based on age, reflecting shared as well as unshared variance.

b. EF performance would be negatively related to parent- and screener-reports regarding behavioral concerns on other screening measures.

c. Exploratory analyses were planned to examine whether the relationship between EF and school readiness varies as a function of age, gender, and socio-economic status (SES). There were no specific hypotheses related to these analyses.

The fourth aim of this study assessed the predictive validity of EF compared to the other screening measures for academic and behavioral outcomes. The following hypotheses were tested:

a. EF performance during screening would be a significant predictor of beginning of kindergarten literacy and numeracy.

b. EF performance would be a significant predictor of kindergarten progress in these domains as well, predicting growth across the school year.

c. Developmental readiness would be a significant predictor of both beginning and end of kindergarten literacy and numeracy

d. Developmental readiness would also predict growth in these areas during the kindergarten year.

e. Developmental readiness was expected to be a significant predictor of end of kindergarten achievement in both domains, above and beyond beginning of kindergarten performance.

f. Considered in the same model, both EF and developmental readiness would each have unique predictive value for literacy and numeracy at the beginning of kindergarten and academic growth across the year

g. EF, as well as the district screening measures for behavior problems, would be moderately related to classroom behavioral problems in the expected directions. There was no specific hypothesis regarding whether performance on the developmental screener would be related to classroom behavior.

h. EF, as well as both measures that the district uses to screen for behavior problems, would have unique predictive value for kindergarten behavior problems.

i. Exploratory analyses was conducted to assess whether any of the measures administered during screening would be predictive of first grade behavior problems, beyond what is predicted by kindergarten behavior. There were no specific hypotheses regarding the longer-term predictive value of EF and current screening tools for behavior problems.

Method

Participants

The current study is part of a larger screening project (N = 606) examining executive functioning in early childhood screening for English, Spanish, Somali, and Hmong speaking children in the MPS district. Because two of the EF measures (DCCS-Dext, and Flanker-Dext) were only available in English at the time of the screening, this study focused on a subsample of participants (N = 461) whose primary language was English. This sample consisted of 461 children (54% female), who were aged 36 to 71 months (M = 54.61, SD = 9.53) at the time of assessment, and their parents. The MPS district is racially and ethnically diverse, as was the sample that participated (34% Black/African American, 44% White/Caucasian, 6% American Indian, 4% Asian, 11% biracial/multiracial; 7% also identified as Hispanic). Children were screened at community locations, including emergency housing (shelters) for families experiencing homelessness or domestic violence. For the study sample, 10% of the children were currently residing in emergency housing. The participation rate of eligible families for the overarching study was 91%.

Procedures

Data for the initial part of the study were previously collected across two summers (July – August 2012, 2013). Families were recruited upon arriving at MPS early childhood screening. Parents who consented gave permission for children to complete the measures of EF during screening as well as permission for the researchers to have access to data collected at screening (e.g., performance on screening measures, demographic information, etc.). Parents also gave permission for the researchers to gather administrative school data about the participants through the third grade, when children complete statewide benchmark tests of achievement. Administrative data available from school records included kindergarten and first-grade records of achievement, behavior, attendance, and special education status for any children still attending school within the district.

After completing procedures for consent, families participated in early childhood screening as usual. Children worked individually with an assessor who administered a developmental screener, provided ratings on observed child behavior, and conducted a brief health screening (e.g., height, weight, vision, hearing). Upon completing screening procedures with the child, the assessor administered a brief tabletop measure of EF. While the children were assessed, parents completed demographic questionnaires and parent report measures of child functioning. At the end of screening, parents typically met with a nurse to review the results and discuss any concerns. At that time, children completed the additional computerized EF measures with either the same assessor who conducted the screening or another member of the research staff.

In the MPS district, children in kindergarten complete academic assessments in September and May. These assessments, which focus on literacy and mathematics skills, are administered by MPS staff, often retired teachers, and not the child's classroom teacher. In addition, at three points during the school year (fall, winter, spring), children's classroom teachers are asked to complete behavioral ratings for children. For any child whose fall score is significantly elevated, indicating behavioral problems, teachers are asked to re-evaluate behavior on the same measure in the winter and spring to monitor progress. However, if there are no concerns for child behavior in the fall, the teacher does not need to repeat the measure. All data on kindergarten and first grade academic outcomes and classroom behavior were obtained through the school district. It was not possible to get follow-up data for all children, which is discussed in more detail below (see Missing Data).

Measures

Demographic

Parents reported key demographic information including children's age at screening, gender, and race/ethnicity. Additionally, parents provided the address to their home residence or it was noted whether a child was screened in an emergency housing and should be considered HHM. The school district does not request information about income or parent education. Therefore, a proxy for income was created with participant addresses using Geocoding (Social Explorer, 2014). This program uses census data to estimate average household income in a given neighborhood; this study utilized the median household income for the census tract (observed range = $13,511 – $125,461). It was not possible to use geocoding to approximate the incomes of children screened in shelter since they do not have a permanent address. Thus, in some analyses socioeconomic status (SES) is represented by geocodes and in others SES is a categorical variable in which children were grouped in low (HHM or estimated yearly income below $37,353), medium (estimated yearly income from $37,353 - $56,066), and high (estimated yearly income above $57,022) groups, in order to include the HHM children. Cutoff points for these categories were determined such that there were approximately equal groups in each category. All analyses will denote whether SES is being used as a categorical variable (low, medium, high) or continuous variable based on geocodes.

Early Childhood Screening

MPS screening procedures involved both child performance and parent-report measures. Children were administered the Minneapolis Preschool Screening Instrument – Revised (MPSI-R; MPSI-R Technical Manual, 2007), which assesses developmental skills that are relevant to later school functioning, including fine motor, language, concepts/cognitive, literacy, and gross motor skills. This measure shows good concurrent validity with other screening tools and has established predictive validity for children in this district. There are established cutoff scores based on children's age for needing referral, needing to be rescreened at a future date, and passing. The English version of the measure consists of 64 dichotomously scored items for 4- and 5-year-old children and 48 items for 3-year-old children (literacy subscale not administered to this age group). Because raw scores, as opposed to cut-off scores, were used for the purposes of this study, the different versions of the measure were equated.

The MPSI-R also has a brief measure of administrator-rated observed child behavior (MPSI-R Behavior; MPSI-R Technical Manual, 2007). After completing the developmental skills portion of the screener, the staff member administering the screener rated the child's behavior on 10 items (alpha = .89). The scale contains items that pertain to child behavior broadly, including items on child impulsivity, activity levels, understanding of directions, and participation. Possible scores on each item range from 0 to 2, with higher scores reflecting more problem behaviors.

Parents completed the Ages and Stages – Social Emotional (ASQ-SE; Squires, Bricker, & Twombly, 2002), which is a validated, non-diagnostic screening tool used to assess the social-emotional competence and problem areas of children ranging from 3 to

66 months of age. Parents in this study completed the 36-month, 48-month, or 60-month-old versions based on their child's age. Depending on the version, the questionnaires range from having 34 to 36 items. Higher scores reflect more behavioral problems and parental concerns. Referral for diagnostic screening and support for parental monitoring is recommended for children whose scores are above specific cutoff scores. Because raw scores were used for this measure and there was slight variation in possible scores based on version administered, these scores were also equated.

Executive Function

Children completed a battery of three EF tasks, consisting of one tabletop measure and two computerized measures that were modified versions of the NIH Toolbox EF measures and include developmental extensions.

On the Peg-Tapping task (Diamond & Taylor, 1996), which provides a measure of inhibitory control, the administrator and child passed a wooden dowel back and forth. Children were told tap the dowel once when the administrator taps twice and to tap twice when the administrator taps once. After practicing the task, the administrator presented the child with 16 counterbalanced trials, in which the child was asked to follow the tapping rules and inhibit their natural tendency to copy the administrator. Diamond, Prevor, Callender, & Druin (1997) discuss the neuropsychological underpinnings of the task. Multiple studies have demonstrated that this measure works well even with children from low-income (e.g., Blair & Razza, 2007) and high-risk backgrounds (e.g., Masten et al., 2012; Obradovic et al., 2010).

To assess cognitive flexibility, children were administered the NIH Toolbox Dimensional Change Card Sort with Developmental Extension (DCCS-Dext; Carlson et al., 2015, See Appendix A for more detail). This measure was based on an earlier version of the task now known as the Minnesota Executive Function Measure (Carlson & Zelazo, 2014). As noted above, it was modified to lower the floor of the DCCS task (Zelazo, 2006). Children are first asked to sort pictures on a computer screen by shape for several trials. They then switch to sorting by color. If children succeed on the first rule switch, they advance to a more challenging version of the task involving more frequent alteration between rules (color or shape switching). However, if children have difficulty with the initial rule switches, they drop down to easier versions that involve sorting simpler objects with less cognitive interference until they reach a baseline. Scores on this task range from -5 to 10, with scores from -5 to 5 reflecting the child's accuracy. Negative scores indicate that the child required the developmental extension. Children who had a high accuracy percentage (above 80%) were given additional points for reaction time. Thus, some children have scores higher than 5.

The second computerized EF task, the NIH Toolbox Flanker with Developmental Extension (Flanker-Dext; Anderson et al., 2015; See Appendix B for more detail) provided a measure of selective attention and resistance to interference from distracting stimuli (Rueda et al., 2004). During this computerized task, the child is presented with a row of five fish and told to selectively attend to the middle fish. Children are told that only the middle fish is hungry and are instructed to feed the middle fish by touching the arrow on the screen that matches the way that the middle fish is pointing. The flankers, which serve as distractors, can be pointing in the same direction (congruent) or opposite direction (incongruent) from the middle fish. The standard NIH Toolbox version of the Flanker task is initially administered to children. If children do not pass either the practice or the basic level, the computer program automatically moves into the developmental extension, which is less challenging. This module simplifies the task for children by including levels in which the middle fish is made larger and distractors are separated by increased distance to the target fish or made a different color. Scores on this task range from -5 to 10, with scores from -5 to 5 reflecting the child's accuracy. Negative scores indicate that the child required the developmental extension. Children who had a high accuracy percentage (above 80%) were given addition points for reaction time. Therefore, some children have scores higher than 5.

An EF composite was developed for use in the correlational tables. Z-scores for each of the three EF tasks were created and then averaged (alpha = .77).

Academic Achievement

During kindergarten in the MPS district, children are assessed in the fall and spring on literacy and numeracy skills. Children are administered the Beginning of Kindergarten Assessment (BKA) during the fall to inform instructional needs and the End of Kindergarten Assessment (EKA) in the spring to document growth (Minneapolis Kindergarten Assessments, 2007). The BKA and EKA measures differ in timing of assessment but largely cover the same content.

Although children completed a broader measure of literacy, this study utilizes only the phonemic awareness, language, and alphabetic principles subscales at the recommendation of the school district. Children's BKA literacy and EKA literacy scores were the mean of the z-scores for each subscale (alpha = .81 for both BKA literacy and EKA literacy). The areas of numeracy assessed included: counting forward, counting backwards, ordering numbers, identifying numbers, identifying numbers that come before and after others, and calculating sums and differences. The district identified that all subscales were suitable for use. Therefore, BKA numeracy and EKA numeracy total scores were used. The total scores were converted to z-scores to be consistent with BKA and EKA literacy.

Classroom Behavior

Kindergarten and first-grade classroom teachers rated children on a brief, 12-item measure that assesses behavior across three areas, including: classroom behaviors (e.g., attends to classroom activities, follows rules), externalizing behaviors (e.g., uses appropriate physical behavior, respects materials and property), and socialization (e.g., handles change appropriately, builds positive peer relationships). Response choices ranged from 1 (almost always) to 5 (rarely). Although the measure is written such that there are three overarching domains, Exploratory Factor Analysis (EFA) of the item-level responses indicated that there was a single underlying factor. Therefore, a raw score composite is used (alpha = .95 for kindergarten, .94 for first grade). Given that most children do not have teacher ratings for winter or spring, only kindergarten and first-grade ratings from the fall were used.

Missing Data

There was complete data (N = 461) for child sex, age, HHM status, total MPSI-R scores (including subscales: fine motor, language, and cognitive), and MPSI-Behavior. All other variables had some amount of missing data. Most variables were considered Missing At Random (MAR) and were imputed using MPlus Version 6 (Muthén & Muthén, 1998-2011) to create 20 imputed datasets. MAR assumption was not met for the literacy subscale of the MPSI-R. The decision to still impute this variable is discussed below. The only variables with missing data that were not imputed were demographic variables, specifically minority status and SES, as well as first-grade behavior. The decision to not impute first-grade behavior was based on the fact that many children in this study have not yet reached first grade so more of this data will become available over time. Additionally, due to the extent of missing data, the overall imputation would not run under the specific coverage requirements. Analyses were conducted using both the original and imputed datasets. Results of the imputed dataset are reported. Any differences in findings are briefly summarized at the end of the results section.

In regards to demographic variables, there was missing data for minority status and SES. Minority status (N = 457) was missing for some children because parents opted to not complete the questions on race and ethnic background. Because SES (N = 398) was approximated based on the child's address, it was missing for any child whose parents did not provide a permanent address to the researchers (N = 19) or whose family was residing in a shelter (N = 44) at the time of screening. Subsequently, when SES was made into a categorical variable (low, medium, high), the 19 children whose addresses were not available continued to be missing.

There was minimal missing data for the district's screening measures. The only measure that had true missing data was the ASQ-SE (N = 458). All parents provided some information; however, when completing the measure several parents missed entire pages

(N = 6), which subsequently invalidated the overall score. Although there was no missing data related to the district's developmental screener, the MPSI-R, the literacy subscale is only administered to 4- and 5-year-old children and is not part of the 3-year-old version of the task. For the purposes of usability of the measure for analyses in this study, this subscale was imputed for 3-year-old children, even though it was not missing at random. Given that this could bias results, analyses were re-run removing literacy altogether to assess the robustness of findings. Findings were consistent whether imputed literacy was used or whether literacy was removed from models entirely.

There was missing data for the EF measures, Peg-tapping (N = 444), DCCS-Dext (N = 430), and Flanker-Dext (N = 417), for a variety reasons. Most common reasons included logistical barriers, such as insufficient time (e.g., the family needed to leave before the tasks could be administered) or computer malfunctioning that precluded administration of the tasks. Given that children were administered the computerized tasks while parents met with a nurse for feedback, on rare occasion the nurse felt the need to observe a child to better understand parent concerns, which was prioritized over participation in the study.

Variables with the most notable proportion of missing data included the kindergarten and first-grade school outcome variables. BKA Literacy (N = 242) and EKA Literacy (N = 205) were available for approximately half of the sample. Children entering kindergarten in fall 2015 were not administered the BKA Literacy due to a change in district testing policy for just that specific year. EKA Literacy was unavailable for this same group of children, as they had only recently finished the end of kindergarten testing at the time of this analysis and this data was not yet available from the district. BKA Numeracy (N = 169) and EKA Numeracy (N = 153) were available for children entering kindergarten in 2012 and 2013; however, in 2014 the district began using a different measure, the Concepts of Mathematics (COM) for assessing numeracy. There was COM data on 94 children who had taken the new measure in the fall and 54 who had taken it in the spring (more data missing in the spring, again, due to the most recent year not yet being available from the district). Although this variable was not directly used in the current study, it was included in the dataset to inform the multiple imputation.

Data Analytic Plan

Results are presented in four sections, consistent with the overarching aims of the study. The specific hypotheses and the corresponding analyses are described below. Note that descriptive statistics (mean and standard deviations of all measures) are presented in Table 1.

Aim 1

The first aim of this study was to assess the range of abilities captured by the newly modified measures of EF, DCCS-Dext and Flanker-Dext. It was hypothesized that the developmental extension of these measures would capture a wider range of abilities, with no observable floor or ceiling effects. Further, it was expected that many children would require the development extension, especially children from higher-risk backgrounds, such as HHM. To assess this, descriptive information on the number of children who needed the developmental extensions of the DCCS-Dext and Flanker-Dext was provided. This information was also provided separately for the subsample of children who were HHM. Information on the range of observed scores and scatter-plots were included to identify whether there were observable ceiling or floor effects.

Aim 2

As part of the second aim, convergent validity of the two new EF tasks was assessed. It was hypothesized that the EF tasks would be related to demographic factors known to be related to EF, such as age, SES, and risk status (HHM). This was assessed through analysis of bivariate correlations and mean difference tests, as appropriate. Further, it was hypothesized that the two new EF tasks would be positively and moderately related to each other as well as another widely-used measure of EF for this age (Peg-Tapping). Bivariate correlations of the EF tasks with one another, controlling for age, were observed.

Aim 3

The third aim of this study involved assessing the construct validity of EF as an indicator of developmental readiness. It was hypothesized that performance on the EF tasks and developmental screener would be highly related yet distinct, demonstrating both shared and unshared variance. It was also thought that EF would be related to other measures at screening, including parent report (ASQ-SE) and screener-observed behavioral concerns (MSPI-R Behavior). This was assessed by observing bivariate correlations, including correlations in which age is controlled. Although age is expected to be related to both EF and screening measures, for there to be good evidence for convergent validity correlations should continue to be moderate in strength after controlling for age.

To further assess the extent to which EF and developmental readiness (performance on the developmental screener) are related, Confirmatory Factor Analysis (CFA) was conducted using EF and developmental readiness latent factors. The indicators for the EF latent factor were the three measures of EF. The indicators for the developmental readiness latent factor consisted of the fine motor, cognitive, language, and literacy subscales measured by the district's developmental screener (MPSI-R). Exploratory analyses were conducted to assess the invariance of the relationship between EF and developmental readiness with respect to age, gender, SES (low, medium, high), and minority status.

Aim 4

The fourth aim involved assessing the predictive validity of EF for academic and behavioral outcomes compared to the other measures administered during screening. It was hypothesized that EF would be a significant predictor of beginning of kindergarten literacy and numeracy. This was first assessed through bivariate correlations of EF and the academic measures. Then Structural Equation Modeling (SEM) was conducted in which the EF latent factor was the only predictor for academic outcomes. In order to assess whether the EF latent factor was predictive of academic growth across the year, direct effects of EF predicting end of kindergarten performance (above and beyond beginning of kindergarten academic performance) were observed. Given that it was also hypothesized that developmental readiness would be a significant predictor of both beginning of kindergarten literacy and numeracy, as well as predict growth across the year, this same set of SEM analyses were conducted for developmental readiness. First bivariate correlations between performance on the developmental screener and academic outcomes were assessed. Then SEM analyses were conducted in which the developmental readiness latent factor was the only predictor for academic outcomes. The predictive validity of developmental readiness for beginning of kindergarten academic performance as well as growth across the school year was observed.

Next, it was hypothesized that both EF and developmental readiness would have unique predictive value for beginning of kindergarten performance as well as growth across the year. The SEM same models as described above were assessed, this time with EF and developmental readiness in the same model.

Finally, it was hypothesized that EF would have unique predictive value for kindergarten classroom behavior problems beyond current screening measures. Bivariate correlations of EF and the other screening measures with kindergarten and first grade behavior were reviewed. In the first SEM analysis, the EF and developmental readiness latent factors as well as the two measures the district uses for screening for possible social-emotional (ASQ-SE) and behavioral problems (MPSI-R Behavior) were included as predictors for kindergarten classroom behaviors. Similar analyses were conducted for predicting first grade behavior to assess the longer-term predictive value of these measures. Again, the EF and developmental readiness latent factors and the other behavioral screening measures were included as predictors for kindergarten and first grade classroom behavior; kindergarten classroom behavior was also used as a predictor in the model predicting first grade behavior.

Results

Need for Developmental Extensions and Range of Abilities Captured (Aim 1)

The developmental extension was provided to 56 percent of children administered the Flanker-Dext and 51 percent of children administered the DCCS-Dext. A subsample of children in this study were screened in emergency homeless shelters (N = 44). As hypothesized, these children demonstrated an even higher need for the developmental extensions. On the DCCS-Dext and Flanker-Dext, 68 percent and 81 percent, respectively, were administered the developmental extension. Observed scores on the DCCS-Dext range from -5.00 to 9.31; only six children scored -5.00, which is the lowest possible score. Observed scores on the Flanker-Dext range from -4.63 to 8.07, indicating that there were no floor effects. Scatterplots of DCCS-Dext and Flanker-Dext scores as they relate to age are provided in Figures 1 and 2. These figures demonstrate that inclusion of the developmental extensions captured a considerably wider range of variability.

Convergent Validity for the Newly Modified EF Measures (Aim 2)

Bivariate correlations for the key demographic variables, EF tasks, and screening measures are presented in Table 1 and Table 2 (controlling for age). Bivariate correlations indicated that each of the three EF tasks was positively correlated with age and SES (approximate median household income). Correlations between EF and age were positive and moderate to strong in strength (range from $r = .45$ to $r = .61$), while correlations between EF and SES were positive yet weak (range from $r = .13$ to $r = .14$). An independent samples *t*-test was used to assess whether there were mean differences in EF scores between children assessed in the community and those residing in an emergency homeless shelter. Levene's F-test indicated that it was appropriate to assume homogeneity of variances for the DCCS-Dext but not for Peg-tapping or Flanker-Dext. Results indicated that HHM status had a statistically significant effect on children's mean EF performance for all EF tasks [Peg-Tapping: $t(58.16) = 4.29$, $p < .001$; DCCS-Dext: $t(428) = 2.95$, $p = .003$; Flanker-Dext: $t(62.30) = 4.72$, $p < .001$], with HHM children performing worse. Bivariate correlations indicated that the three EF tasks were strongly related to one another (range of $r = .61$ to $r = .68$, raw scores). The strength of the correlations continued to be strong even when controlling for age ($r = .47$ to .56, residuals after controlling for age).

Convergent Validity for Executive Function (EF) as a Measure of Developmental Readiness (Aim 3)

Bivariate correlations indicated that the three EF tasks were all highly related to the developmental screener administered by the district (range $r = .62$ to .67). Using aged-controlled residuals, correlations for each EF task with the developmental screener ranged from $r = .49$ to $r = .56$. This indicates both shared and unshared variance between these two constructs, and that the shared variance is not based on age alone. Correlations between the EF tasks and other measures at screening, including screener-rated observed behavior and parent-rated social-emotional problems, were also assessed. With respect to screener-rated behavior problems, correlations with each EF task ranged from $r = -.33$ to $r = -.42$; age-controlled correlations were slightly weaker ($r = -.23$ to $r = -.32$). Correlations between the EF tasks and the parent-report measure ranged from $r = -.26$ to $r = -.30$; controlling for age had negligible impact on the strength of these relationships ($r = -.26$ to $r = -.32$).

The next step in assessing convergent validity involved using CFA. It was not possible to use CFA to assess the fit of the EF latent factor independently, given that this model would be just-identified. Thus, the first CFA fit was a two-factor model with EF and developmental readiness latent factors (see Figure 3). The indicators for the EF latent factor were the three measures of EF. The indicators for the developmental readiness latent factor consisted of the fine motor, cognitive, language, and literacy subscales measured by the district's developmental screener.

Of note, the developmental screener includes an additional subscale – gross motor – which was not included in these analyses. Exploratory factor analysis (EFA) of the developmental screener subscales indicated that fine motor, language, cognitive, and literacy had acceptable loadings on a single factor, while gross-motor primarily loaded on a second factor. Therefore, for the purposes of this study, gross motor was excluded as an indicator in the model.

In order to control for age the indicators were the residuals, with age removed, of the three EF tasks and the developmental screener subscales. The indicators in this model had varying relationships with age. Specifically, each of the three EF tasks as well as literacy subscale from the developmental screener increased linearly with age, while the relationship between language, cognitive, and fine motor with age was quadratic. Thus, residuals controlling for either age or age-squared as appropriate were used as indicators in the model. Fit indices for the model were as follows: $\chi^2(13) = 32.722$, $p = .002$; comparative fit index = 0.986; Tucker-Lewis index = 0.977; root mean square error of approximation = 0.057; standardized root mean square residual = 0.027. See Table 5 and Figure 3 for factor loadings.

Next, the invariance of the covariance between the EF and developmental readiness latent factors was assessed. It should be noted that this analysis focused only on the relationship between the EF and developmental readiness latent factors, and did not assess invariance of each pathway in the model. Specifically, chi-square difference tests were conducted to assess whether this pathway was invariant based on child's age in years at screening (3, 4, or 5 years old), sex (male, female), minority status (minority, non-minority), and SES (low, medium, high). For each category, a multi-group model was fit in which the factor loadings, intercepts, and residual variances were constrained for each indicator, the variances were constrained for each factor, and the covariance between EF and developmental readiness was constrained. Then, a second multi-group model was fit in which the same parameters were constrained, except the covariance between the EF and developmental readiness latent factors, which was freely estimated.

For age, the chi-square difference test rejected the null hypothesis [χ^2 (2) = 8.556, p = .014], suggesting that invariance for this path cannot be assumed. The covariance between EF and developmental readiness was stronger for older children. For 3-year-old children, the covariance was .362 (r = .625); for 4-year-old children it was .450 (r = .846); and for 5-year-old children it was .489 (r = .846). Similar analyses and chi-square difference tests were conducted to assess invariance for sex [χ^2 (1), = .226, p = .606], minority status [χ^2 (1) = .283, p = .595] and SES [χ^2 (2) = .441, p = .802], all of which allowed for retaining the null hypothesis and supported the assumption of invariance.

Predictive Validity of Executive Function, Developmental Readiness, and Other Screening Measures for Behavioral and Academic Outcomes (Aim 4)

Bivariate correlations between EF measures, the developmental screener, and kindergarten academic outcomes are summarized in Tables 3 and 4. Overall, after controlling for age at screening, the EF tasks were moderately correlated with beginning of kindergarten literacy (range of $r = .37$ to $r = .42$) and numeracy ($r = .31$ to $r = .43$). Age-controlled correlations were similar in strength for EF and end of kindergarten literacy (range $r = .39$ to $r = .44$) and numeracy ($r = .38$ to $r = .39$). Age-controlled correlations between the developmental screener and beginning of kindergarten literacy ($r = .66$) and numeracy ($r = .59$), as well as end of kindergarten literacy ($r = .61$) and numeracy ($r = .58$), were somewhat stronger.

SEM was used to assess the predictive value of the EF latent factor for beginning of kindergarten achievement and to assess whether EF had significant predictive value for end of kindergarten achievement, beyond children's beginning of kindergarten performance (Figure 4). Age at screening, gender, minority status, and SES were included as control variables for each academic outcome. The only significant path was age at screening as a predictor of end of kindergarten numeracy; therefore this variable was left in the model as a control [$\beta = -.243(.072)$, $p = .001$; B = -.312(.093), $p = .001$]. The model showed good overall fit: χ^2 (11) = 14.043, $p = .231$; comparative fit index = 0.995; Tucker-Lewis index = 0.987; root mean square error of approximation = 0.024; standardized root mean square residual = 0.021. Standardized and unstandardized estimates for each path are summarized in Table 6.

Results indicated that the EF latent factor was a significant predictor for each academic outcome. It was a moderately strong predictor of both literacy and numeracy at the beginning of kindergarten. Although beginning of kindergarten performance was a stronger predictor, EF continued to be a significant predictor for children's end of kindergarten performance. Of note, when EF was the only predictive path in the model for literacy and numeracy end of kindergarten achievement (leaving the control variable but removing beginning of kindergarten literacy and numeracy performances), the predictive strength increased [Literacy: β = 0.584(.053), $p < .001$; B = 0.832(.119), $p < .001$; Numeracy: β = 0.501(.059), $p<.001$; B = 0.815(.105), $p < .001$].

Next, a model was fit assessing the predictive value of the developmental readiness latent factor (Figure 5). Similar to the above approach, this model assessed the predictive value of the developmental factor for beginning of kindergarten performance and end of kindergarten achievement, beyond what was accounted for by beginning of kindergarten performance. Age at screening, gender, minority status, and SES were included as control variables. Again, the only significant path for the control variables was age at screening on end of kindergarten numeracy, so this variable was retained as a control in the model [β = -.241(.071), p = .001; B = -.310(.091), $p<.001$]. Model fit indices are as follows: χ^2 (18) = 59.376, $p < .001$, comparative fit index = 0.889; Tucker-Lewis index = 0.945; root mean square error of approximation = 0.071; standardized root mean square residual = 0.048. Standardized and unstandardized estimates for each path are summarized in Table 7.

Results suggest that the developmental readiness latent construct is a significant predictor for each path. Similar to the model with the EF latent factor, developmental readiness was a moderately strong predictor of the beginning of kindergarten assessment for both literacy and numeracy. Again, with literacy and numeracy beginning of kindergarten assessments and developmental readiness as predictors of end of kindergarten achievement, developmental readiness had significant, although weak, predictive power. The model was re-run in order to assess the predictive strength of developmental readiness without beginning of kindergarten performance. In this model, the predictive strength of developmental readiness for end of kindergarten assessment increased [Literacy: β = 0.635(.046), p < .001; B = 0.585(.066), p < .001; Numeracy: B = 0.612(.070), p<.001; B = 0.644(.081), p < .001].

Finally, the same general approach was used as the above models; however this time, both EF and Developmental Readiness latent factors were allowed to compete in the same model (Figure 6). As in the previous models, age at the time of screening was included as a control variable for all academic outcomes and was a significant predictor of end of kindergarten numeracy [β = -.241(.071), p = .001; B = -.310(.091), p = .001]. No other controls were retained. Fit indices for the model were as follows: χ^2 (40) = 110.438, p < .001, comparative fit index = 0.943; Tucker-Lewis index = 0.905; root mean square error of approximation = 0.062; standardized root mean square residual = 0.044. Standardized and unstandardized estimates for each path are summarized in Table 8.

In this model, the EF latent factor was no longer a significant predictor of any kindergarten academic outcome. Developmental readiness continued to be a significant predictor of moderate strength for literacy and numeracy at the beginning of kindergarten. The only significant predictor of end of kindergarten literacy and math achievement was beginning of kindergarten performance. The model was then run again, removing beginning of kindergarten assessments as a predictor for end of kindergarten achievement in order to assess whether EF and developmental readiness had any significant direct effect on end of kindergarten achievement in this model. Results indicated that EF was a significant predictor of end of kindergarten literacy [β = 0.234(.100), p = .034; B = 0.342(.166), p = .040] but not numeracy [β = 0.163(.113), p =.149; B = 0.270(.189), p = .152]. Developmental readiness was a significant predictor of both literacy [β = 0.455(.104), $p < .001$; B = 0.419(.102), $p < .001$] and numeracy [β = 0.486(.129), p<.001; B = 0.511(.138), $p < .001$].

Note that in each of the above models, the predictive pathways for EF and developmental readiness were assessed for moderating effects based on age at screening (3, 4, or 5 years old), sex, minority status, SES (low, medium, high). There were no significant findings.

To address the predictive value of EF for kindergarten and first-grade classroom behavioral problems, age-controlled bivariate correlations were observed (Table 4). EF measures at screening were negatively correlated with kindergarten (range of r = -.21 to r = -.34) and first-grade (range of r = -.31 to r = -.41) behavioral problems. The strengths of the correlations were comparable to other measures used during screening for behavior problems, including a measure of parent-reported concerns regarding child social and emotional development (ASQ-SE; kindergarten r = .33; first grade r = .37) and screener-observed behavior problems (MPSI-R Behavior; kindergarten r = .32; first grade r = .34).

SEM was used to assess the unique predictive value of each measure for kindergarten classroom behavior (Figure 7). Maximum likelihood estimation with robust standard errors (MLR) was used for these analyses due to kindergarten classroom behavior having a positively skewed distribution. Fit indices of the model were as follows: $\chi^2(28)$ =

79.819, comparative fit index = 0.967; Tucker-Lewis index = 0.948; root mean square error of approximation = 0.061; standardized root mean square residual = 0.031. Standardized and unstandardized estimates for each path are summarized in Table 9. The EF latent factor, parent-reported social-emotional concerns (ASQ-SE), and screener-observed behavior problems (MPSI-R Behavior) were all significant predictors, in the expected directions, of kindergarten classroom behavior problems. EF was a moderately strong predictor, and the strongest predictor in the model.

The next model assessed whether any of the measures administered at screening had unique predictive value for first-grade behavior, above and beyond kindergarten behavior (Figure 8). Again, MLR was used for these analyses due to both kindergarten and first-grade classroom behavior problems having positively skewed distributions. Fit indices were for the model were: χ^2 (33) = 88.994, comparative fit index = 0.996; Tucker-Lewis index = 0.944; root mean square error of approximation = 0.061; standardized root mean square residual = 0.033. Standardized and unstandardized estimates for each path are summarized in Table 10. In this model, only kindergarten classroom behavior problems and the EF latent factor were significant predictors of first-grade behavior problems. Notably, the predictive value of EF continued to be moderately strong, despite the fact that it was measured 1 to 3 years prior to children beginning first grade.

In both SEM models, age at screening (3, 4, or 5 years old), sex, minority status, and SES were included as control variables to assess for having main effects on kindergarten and first-grade behavior. There were no significant findings, so these variables were excluded from the model. Each predictive path was assessed for moderating effects based on these same variables. There were no significant findings.

Summary of Discrepant Findings from the Original (Non-Imputed) Dataset

Follow-up analyses were conducted using the original non-imputed dataset. Overall, findings were similar. For the models in which EF (i.e., the SEM model displayed in Figure 4) was considered independently as a predictor of beginning and end of kindergarten achievement, the direct path to end of kindergarten numeracy was no longer significant [β = .149(.087), *p* = .086; B = .270(.160), *p* = .091]. This was also true in the model for which developmental readiness (i.e., the SEM model displayed in Figure 5) was considered independently; developmental readiness was no longer a significant predictor of end of kindergarten numeracy [β = .142(.093) *p* = .127; B = .164(.111) *p* = .140]. This may be due to decreased power and difficulty detecting weaker effects. The only other notable changes were slight variations in which control variables had significant effects on academic outcomes. This information is available upon request.

Discussion

The current study represents the first known effort to include direct testing of EF during early childhood screening in order to assess the potential value of EF assessment as part of routine screening in a large and diverse, urban school district. Results indicate very promising construct and predictive validity.

Findings indicated that EF performance is strongly related to school readiness defined by a broad developmental screening test and moderately related to other validated measures of behavior related to school adjustment. These findings suggest that EF performance is related to other aspects of development assessed, such as broad cognitive development, language, and fine motor functioning, all of which are known to be important for early school success (Brassard & Boehm, 2007; MPSI-R Technical Manual, 2007; National Research Council, 2008). This finding also provides convergent validity for EF as a measure of developmental readiness and supports previous suggestions that aspects of self-regulation, such as EF, may be foundational skills for school readiness and later academic success (Blair, 2002; Blair & Raver, 2015).

In addition to understanding the relationship between EF and performance on the developmental screener, it was important to assess whether there are particular groups of children for whom this relationship is more or less strong. Findings indicated that the relationship between EF and developmental readiness was invariant based on sex, minority status, and SES. This was not true for age, as the relationship between the two constructs was stronger for older children. While it is possible that these constructs truly overlap less for younger children, it is plausible that younger children show more overall variability in performance and consequently more error is captured by each measure, subsequently weakening the correlation. Although invariance for age cannot be assumed, the correlation for 3-year-old children was still strong and above the .60 correlation coefficient threshold recommended by Brassard and Boehm (2007) for establishing convergent validity.

Predictive validity of the EF measures was also established as a part of this study. Findings were consistent with existing literature suggesting that better EF skills are associated with stronger academic performance (e.g., Bull & Scerif, 2001; Bull et al., 2008; Clark et al., 2010; Masten et al., 2012; Mazzocco & Kover, 2007). Children with better EF at screening had both better initial literacy and numeracy skills at the beginning of kindergarten and made greater gains across the school year.

Although EF at screening was moderately related to literacy and numeracy at the beginning and end of kindergarten, neither it nor the developmental screener were as highly correlated as Brassard and Boehm (2007) suggest would be ideal (.80). As the authors discussed, while very strong correlations are desirable, even good measures are unlikely to be able to account for the multitude of external factors such as parental involvement, teacher skill level, and child attendance that all impact young children's achievement. These findings reinforce the need for brief measures to be used as screeners – to help identify children who need further evaluation – and not be used for informing high-stakes decisions like placement in special education or readiness kindergartens.

Further analyses indicated that while EF accounted for a significant portion of the variance when considered independently, it was not a stronger predictor than the developmental screener nor did it contribute unique predictive value for understanding academic outcomes. Ultimately, when considered together, the variance being predicted by EF was accounted for entirely by the developmental screener. This may in part be due to the overlap in the neurocognitive processes involved in completing both tasks. In other words, engaging in any type of cognitive task involves EF skills (Alexander & Stuss, 2000; Della Sala et al., 1998; Denckla, 1994), and thus, the developmental screener also likely captured these skills to a strong degree. In addition, multiple measures used (MPSI-R, BKA, and EKA) were explicitly designed by the same school district to assess key aspects of learning valued by this district, and clearly the MPSI-R is a successful predictor of these similar achievement measures in kindergarten. Thus, it is not surprising that EF contributes so little unique value to predicting these initial outcome measures of achievement.

EF did, however, have unique predictive value for behavior, compared to the other district screening measures. Findings indicated that EF was a moderately strong predictor of both kindergarten and first-grade behavior problems. Although both of the measures currently used by the MPS district to assess possible behavioral problems were significant predictors of kindergarten behavior problems, EF was the strongest predictor. Further, EF was the only measure to continue to have predictive value for first-grade behavior problems. Despite performance on the EF measures and developmental screener having strong overlap, developmental readiness was not predictive of classroom behavior problems in kindergarten or first grade. This suggests that there is value in the unshared variance between EF and the developmental screener (that is captured by EF) for understanding risk for later behavioral problems.

A separate aim of this study was to assess the psychometric properties of two newly modified measures of the NIH Toolbox tasks. Results demonstrated that the two EF tasks, the DCCS-Dext and Flanker-Dext, captured a considerably wider range of abilities than the original NIH versions of the task. On both tasks, the majority of children required the developmental extension. Although several children still earned the minimum score on DCCS-Dext, floor effects for both tasks were minimal. Further, the need for the developmental extensions of both measures was greater among the subsample of children residing in emergency homeless shelters, suggesting that researchers and educators measuring EF for children from adverse backgrounds must be especially careful about selecting developmentally sensitive tools that have sufficiently low floors.

Findings supported convergent validity for the newly modified measures as well, as both measures were moderately related to each other and a third, widely-used measure of EF, the Peg-Tapping task (Diamond & Taylor, 1996), as well as age and SES. Of note, the relationship between EF and SES may actually be stronger than what is suggested in this study, as geocoding – which was used to approximate SES – was not inclusive of HHM children. Children who were HHM at the time of screening performed significantly poorer on all three EF tasks, building on previous literature supporting that the development of EF is susceptible to early life stress (e.g., DePrince, et al., 2009; Loman et al., 2012; Pears et al., 2010; Pechtel & Pizzagalli, 2011).

Implications

Taken together, the current study demonstrated the substantial value of brief developmental screening for assessing children's risk for early school difficulty and the added value, especially with respect to predicting behavioral problems, of including assessment of EF. Part of the interest in measuring EF stems from research supporting that it is predictive of school outcomes (e.g., Bull & Scerif, 2001; Bull et al., 2008; Clark et al., 2010; Masten et al., 2012; Mazzocco & Kover, 2007) and malleable to intervention (Diamond & Lee, 2011; Diamond & Ling, 2016 for a review). Findings from this study support continuing efforts to try to boost EF in children prior to the start of school, particularly for disadvantaged children or children developing in high-risk contexts who are at risk for lower achievement (Cutuli et al., 2013; Herbers et al., 2012). Interventions for children with EF difficulties could be in the form of providing individual training, EF-focused classrooms, or supporting parenting (e.g., Bodrova & Leong, 2007; Bernier, Carlson, & Whipple, 2010; Bierman et al., 2008; Diamond & Ling, 2016; Raver et al., 2009)

Limitations and Future Directions

The current study had several limitations that may be useful to address in future research. First, prior to EF being used as a screener, it will be critical to develop norms and cut-off scores to indicate whether a child a) passed b) needs to be re-screened or c) needs referral. Next, the sensitivity and specificity would need to be assessed (reviewed in Gredler, 1997). The sensitivity index of a measure represents the probability that a child that is not at risk will be correctly identified as not at risk, while specificity represents the probability that a child that is not at risk will be correctly identified as not at risk. Meisels (1999) recommends that screening measures, at a minimum, have both a sensitivity and specificity index of .80, while Brassard and Boehm (2007) recommend a sensitivity index of .90 and a specificity index of .80.

A second limitation of the current study involves assessment of invariance in the EF and developmental readiness CFA. As indicated above, the only pathway in this model that was assessed for invariance was the covariance between the EF and developmental readiness latent factors. Future research should assess the measurement invariance of the two constructs more broadly, including all paths in the model.

Further, an additional next step that may prove useful is for future studies to have a more multi-faceted measure of child behavioral outcomes, including parent-report of child functioning. The present study was able to assess the predictive value of the measures administered at screening for classroom behavior; however, it is possible that three different measures (ASQ-SE, MPSI-Behavior, and EF) may be more or less valuable depending on the outcome of interest. For example, it is reasonable to believe that EF may be the best predictor of child classroom behavior but that another measure, such as the ASQ-SE, may be a better predictor of child clinical problems.

Lastly, as a word of caution, future research on the value of EF as a screener may be interested in whether it is possible to meaningfully assess EF using a single measure, in order to save time and resources. This approach, however, may have limitations. As Willoughby and Blair (2011) summarize, young children’s performance may vary on tasks in part because of idiosyncratic aspects of each task. Although performance on any given measure of EF may be less stable when considered individually, there is evidence supporting that the underlying EF latent factor is quite reliable (Ettenhofer, Hambrick, & Abeles, 2006; Willoughby & Blair, 2011).

Tables

Table 1. *Intercorrelations of Demographic, Executive Function, and Screening Measures*

Variable	Sex	Age	Minority	SES	HHM	EF-Composite	Peg-tapping	DCCS-Dext
Sex (Female)	--							
Age (Months)	-.01	--						
Minority	-.01	.03	--					
SES	.04	-.06	-.39	--				
HHM	.01	-.08	.26	NA	--			
EF Composite	.04	.59	-.27	.16	-.18	--		
Peg-tapping	-.01	.45	-.23	.13	-.16	.85	--	
DCCS-Dext	.07	.48	-.23	.14	-.14	.88	.62	--
Flanker-Dext	.03	.61	-.25	.14	-.16	.88	.61	.68
MPSI-R Total	.03	.55	-.33	.22	-.29	.75	.62	.67
Fine Motor	.05	.55	-.20	.07	-.22	.64	.53	.58
Language	.10	.45	-.30	.20	-.28	.61	.48	.58
Cognitive	.00	.49	-.34	.20	-.31	.74	.61	.67
Literacy	-.01	.46	-.38	.28	-.28	.65	.54	.57
MPSI-R Beh.	-.07	-.30	.15	-.12	.05	-.43	-.33	-.42
ASQ-SE	-.02	-.09	.31	-.21	.17	-.33	-.26	-.30
Mean	.54	54.61	0.56	55948.35	.10	-.03	8.74	.58
(SD)	-	(9.52)	-	(25822.61)	-	(.87)	(5.64)	(3.41)

Table 1 (continued). *Intercorrelations of Demographic, Executive Function, and Screening Measures*

Variable	Flanker-Dext	MPSI-Total	Fine Motor	Lang.	Cogn.	Lit.	MPSI-Behavior	ASQ-SE
EF Composite								
Peg-tapping								
DCCS-Dext								
Flanker-Dext	--							
MPSI-R Total	.66	--						
Fine Motor	.56	.79	--					
Language	.54	.89	.70	--				
Cognitive	.65	.93	.73	.84	--			
Literacy	.58	.86	.57	.67	.75	--		
MPSI-R Beh.	-.38	-.56	-.45	-.57	-.55	-.39	--	
ASQ-SE	-.29	-.37	-.29	-.38	-.39	-.28	.34	--
Mean	1.53	48.18	5.89	17.74	10.19	10.61	.97	34.77
(SD)	(3.00)	(12.97)	(1.64)	(4.23)	(3.21)	(5.76)	(2.31)	(37.60)

Notes: The dataset for the above variables was multiply imputed with the exception of demographic information. N = 461 except for Minority (N = 457) and SES (N = 398). Bivariate correlations were pooled across 20 datasets.

ASQ-SE and MPSI-R Behavior are scored such that higher scores represent more problem behavior. HHM refers to Homeless Highly Mobile; EF Composite is the z-score average of the three executive functioning measures (Peg-tapping, DCCS-Dext, and Flanker-Dext). MPSI-R total is the raw score sum on this measure. HHM and Minority status are dichotomous variables, scored 0 = No, 1 = Yes. Due to SES being approximated with geocodes, this data is not available for children who were HHM at screening, as they did not have a permanent residence.

Table 2. *Age-Controlled, Intercorrelations of Demographic, Executive Function, and Screening Measures*

Variable	Sex	Minority	SES	HHM	EF-Comp	Peg-tapping	DCCS-Dext	Flanker-Dext	MPSI-Total	Fine Motor	Lang.	Cogn.	Lit.	MPSI-Behavior	ASQ-SE
Sex (Female)	--														
Minority	-.01	--													
SES	.04	-.39	--												
HHM	.01	.26	NA	--											
EF Comp.	.06	-.36	.24	-.16	--										
Peg-tapping	.09	-.28	.18	-.14	.81	--									
DCCS-Dext	.05	-.27	.19	-.12	.85	.51	--								
Flanker-Dext	.06	-.33	.23	-.14	.81	.47	.56	--							
MPSI-R Total	.04	-.42	.30	-.29	.63	.50	.56	.49	--						
Fine Motor	.07	-.25	.12	-.21	.46	.37	.43	.34	.69	--					
Language	.12	-.35	.25	-.27	.48	.35	.46	.36	.86	.61	--				
Cognitive	.01	-.40	.25	-.32	.63	.49	.56	.50	.89	.65	.79	--			
Literacy	.00	-.45	.34	-.27	.52	.42	.44	.42	.81	.43	.58	.68	--		
MPSI-R Beh.	-.08	.17	-.14	.03	-.33	-.23	-.32	-.26	-.49	-.36	-.51	-.47	-.30	--	
ASQ-SE	-.01	.33	-.09	.16	-.35	-.26	-.30	-.32	-.37	-.26	-.34	-.36	-.27	.31	--

Notes: The dataset for the above variables was multiply imputed with the exception of demographic information. N = 461 except for Minority (N = 457) and SES (N = 398). Bivariate correlations were pooled across 20 datasets. The italicized variables are the residuals, after controlling for age.

ASQ-SE and MPSI-R Behavior are scored such that higher scores represent more problem behavior. HHM refers to Homeless Highly Mobile; EF Composite is the z-score average of the three executive functioning measures (Peg-tapping, DCCS-Dext, and Flanker-Dext). MPSI-R total is the raw score sum on this measure. HHM and Minority status are dichotomous variables, scored 0 = No, 1 = Yes. Due to SES being approximated with geocodes, this data is not available for children who were HHM at screening, as they did not have a permanent residence.

Table 3. *Intercorrelations of Demographic, Executive Function, Screening Measures, and School Outcomes*

	BKA Literacy	BKA Numeracy	EKA Literacy	EKA Numeracy	K Behavior	1st Grade Behavior
Child Sex (Female)	.05	-.02	.05	.03	-.09	-.17
Child Age	-.02	.03	.01	.27	.05	.04
Minority	-.33	-.36	-.31	-.29	.21	.21
SES	.24	.23	.18	.15	-.15	-.09
HHM	-.22	-.23	-.19	-.18	.07	.20
EF Composite	.38	.37	.42	.54	-.23	-.33
Peg-tapping	.37	.39	.40	.46	-.17	-.27
DCCS	.32	.28	.35	.46	-.19	-.29
Flanker	.31	.30	.36	.47	-.24	-.30
MPSI-R Total	.54	.51	.52	.64	-.21	-.25
Fine Motor	.30	.25	.32	.52	-.11	-.10
Language	.50	.43	.50	.57	-.22	-.20
Cognitive	.52	.50	.52	.62	-.22	-.30
Lit	.61	.57	.52	.58	-.26	-.22
MPSI-R Behavior	-.32	-.25	-.29	-.50	.29	.32
ASQ-SE	-.34	-.33	-.33	-.37	.31	.23
Mean	-.02	.00	-.04	.00	16.48	17.38
(SD)	(.88)	(1.00)	(.87)	(1.00)	(7.56)	(8.03)

Notes: The dataset for the above variables was multiply imputed with the exception of demographic information and First Grade Behavior. N = 461 except for Minority (N = 457), SES (N = 398) and 1st Grade Behavior (N = 165).

ASQ-SE, MPSI-R Behavior, Kindergarten Behavior, and First Grade Behavior are scored such that higher scores represent more problem behavior. HHM refers to Homeless Highly Mobile; BKA refers to Beginning of Kindergarten Assessment; EKA refers to End of Kindergarten Assessment; EF Composite is the z-score average of the three executive functioning measures (Peg-tapping, DCCS-Dext, and Flanker-Dext). MPSI-R total is the raw score sum on this measure. HHM and Minority status are dichotomous variables, scored 0 = No, 1 = Yes. Due to SES being approximated with geocodes, this data is not available for children who were HHM at screening, as they did not have a permanent residence.

Table 4. *Age-controlled Intercorrelations of Executive Function, Screening Measures, and School Outcomes*

	BKA Literacy	BKA Numeracy	EKA Literacy	EKA Numeracy	K Behavior	1st Grade Behavior
EF Composite	.48	.44	.51	.57	-.32	-.43
Peg-tapping	.42	.43	.44	.38	-.21	-.31
DCCS	.37	.31	.39	.39	-.25	-.35
Flanker	.40	.36	.44	.38	-.34	-.41
MPSI-R Total	.66	.59	.61	.58	-.28	-.32
Fine Motor	.36	.28	.37	.44	-.15	-.14
Language	.56	.47	.56	.50	-.24	-.24
Cognitive	.60	.56	.58	.56	-.28	-.35
Lit	.67	.64	.55	.51	-.25	-.28
MPSI-R Behavior	-.34	-.25	-.29	-.43	.32	.34
ASQ-SE	-.34	-.32	-.32	-.34	.33	.37

Notes: The dataset for the above variables was multiply imputed with the exception of demographic information and 1st Grade Behavior. N = 461 except for Minority (N = 457), SES (N = 398) and 1st Grade Behavior (N = 165). Bivariate correlations were pooled across 20 datasets. All variable assessed at the time of screening (listed in the left hand column) are the residuals, after controlling for age.

ASQ-SE, MPSI-R Behavior, Kindergarten Behavior, and First Grade Behavior are scored such that higher scores represent more problem behavior. BKA refers to Beginning of Kindergarten Assessment; EKA refers to End of Kindergarten Assessment; EF Composite is the z-score average of the three executive functioning measures (Peg-tapping, DCCS-Dext, and Flanker-Dext). MPSI-R total is the raw score sum on this measure.

Table 5. *Factor Loadings of the Executive Function / Developmental Readiness Measurement Model*

Factor and Indicator	Standardized Estimate (Std. Error)	*p*	Unstandardized Estimate (Std. Error)	*p*
Executive Function				
Peg-tapping	.662(.034)	<.001	.848(.068)	<.001
DCCS	.785(.027)	<.001	1.000(.000)	--
Flanker	.707(.031)	<.001	.801(.059)	<.001
Developmental Readiness				
Fine Motor	.684(.027)	<.001	1.000(.000)	--
Language	.822(.018)	<.001	3.275(.201)	<.001
Cognitive	.956(.012)	<.001	2.815(.161)	<.001
Lit	.709(.033)	<.001	3.800(.318)	<.001
Factor Correlation				
EF with DR	.748(.031)	<.001	.496(.053)	<.001

Notes: Dataset was multiply imputed (N = 461). Factor loadings were pooled across 20 datasets. All indicators are age-controlled residuals.

Model fit indices: $\chi2(13) = 32.722$, $p = .002$; comparative fit index = 0.986; Tucker-Lewis index = 0.977; root mean square error of approximation = 0.057; standardized root mean square residual = 0.027.

Table 6. Structural Equation Modeling of Executive Function as a Predictor of Kindergarten Literacy and Numeracy

Path	Standardized Estimate (Std. Error)	*p*	Unstandardized Estimate (Std. Error)	*p*
EF → BKA Literacy	.551(.052)	<.001	.790(.105)	<.001
EF → BKA Numeracy	.501(.059)	<.001	.815(.105)	<.001
EF → EKA Literacy	.219(.054)	<.001	.312(.085)	<.001
EF → EKA Numeracy	.192(.075)	.010	.314(.126)	.013
BKA Literacy → EKA Literacy	.706(.051)	<.001	.700(.062)	<.001
BKA Literacy → EKA Numeracy	.264(.080)	.001	.300(.092)	<.001
BKA Numeracy → EKA Literacy	.706(.051)	<.001	-.041(.053)	.438
BKA Numeracy → EKA Numeracy	.390(.080)	<.001	.391(.080)	<.001
BKA Literacy with BKA Numeracy	.591(.057)	<.001	.370(.053)	<.001
EKA Literacy with EKA Numeracy	.330(.085)	<.001	.106(.031)	<.001

Note: Dataset was multiply imputed (N = 461). Above are the standardized and unstandardized estimates, pooled across 20 datasets. EF refers to Executive Functioning; BKA refers to Beginning of Kindergarten Assessment; and EKA refers to End of Kindergarten Assessment.

Model fit indices: $\chi 2$ (11) = 14.043, p = .231; comparative fit index = 0.995; Tucker-Lewis index = 0.987; root mean square error of approximation = 0.024; standardized root mean square residual = 0.021.

Control variables: Gender, minority status, SES (geocode), and years from screening until starting kindergarten were assessed as control variables. Only years until starting kindergarten was a significant predictor for EKA numeracy [β = -.243(.072), p = .001; B = -.312(.093), p = .001], so it was retained in the final model. All other control variables were removed.

Moderating variables: Gender, minority status, SES (geocode), and years from screening until starting kindergarten were assessed for having a moderating effect on the predictive strength of EF for outcome variables. There were no significant effects and therefore these paths were not retained in the final model.

Table 7. *Structural Equation Modeling of Developmental Readiness as a Predictor of Kindergarten Literacy and Numeracy*

Path	Standardized Estimate (Std. Error)	*p*	Unstandardized Estimate (Std. Error)	*p*
DR → BKA Literacy	.668(.040)	<.001	.620(.061)	<.001
DR → BKA Numeracy	.603(.069)	<.001	.636(.084)	<.001
DR → EKA Literacy	.205(.069)	.003	.189(.064)	.003
DR → EKA Numeracy	.240(.107)	.025	.252(.113)	.025
BKA Literacy → EKA Literacy	.695(.057)	<.001	.690(.064)	<.001
BKA Literacy → EKA Numeracy	.227(.092)	.014	.258(.105)	.014
BKA Numeracy → EKA Literacy	-.055(.074)	.461	-.048(.065)	.457
BKA Numeracy → EKA Numeracy	.366(.076)	<.001	.366(.077)	<.001
BKA Literacy with BKA Numeracy	.502(.072)	<.001	.258(.052)	<.001
EKA Literacy with EKA Numeracy	.339(.080)	<.001	.110(.030)	<.001

Note: Dataset was multiply imputed (N = 461). Above are the standardized and unstandardized estimates, pooled across 20 datasets. DR refers to Developmental Readiness; BKA refers to Beginning of Kindergarten Assessment; and EKA refers to End of Kindergarten Assessment.

Model fit indices: $\chi2$ (18) = 59.376, p < .001; comparative fit index = 0.889; Tucker-Lewis index = 0.945; root mean square error of approximation = 0.071; standardized root mean square residual = 0.048.

Control variables: Gender, minority status, SES (geocode), and years from screening until starting kindergarten were assessed as control variables. Only years until starting kindergarten was a significant predictor for EKA numeracy [β = -.241(.071), *p* = .001; B = -.310(.091), *p*<.001], so it was retained in the final model. All other control variables were removed.

Moderating variables: Gender, minority status, SES (geocode), and years from screening until starting kindergarten were assessed for having a moderating effect on the predictive strength of DR for outcome variables. There were no significant effects and therefore these paths were not retained in the final model.

Table 8. *Structural Equation Modeling of Executive Function and Developmental Readiness as Predictors of Kindergarten Literacy and Numeracy*

Path	Standardized Estimate (Std. Error)	*p*	Unstandardized Estimate (Std. Error)	*p*
EF → BKA Literacy	.102(.087)	.241	.151(.129)	.243
EF → BKA Numeracy	.088(.129)	.495	.144(.213)	.500
EF → EKA Literacy	.167(.087)	.053	.244(.130)	.060
EF → EKA Numeracy	.110(.097)	.256	.183(.162)	.259
DR → BKA Literacy	.586(.077)	<.001	.544(.082)	<.001
DR → BKA Numeracy	.536(.142)	<.001	.564(.154)	<.001
DR → EKA Literacy	.086(.105)	.412	.079(.096)	.411
DR → EKA Numeracy	.160(.137)	.245	.167(.145)	.250
EF with DR	.754(.031)	<.001	.429(.050)	<.001
BKA Literacy → EKA Literacy	.686(.054)	<.001	.680(.061)	<.001
BKA Literacy → EKA Numeracy	.224(.091)	.014	.224(.091)	.014
BKA Numeracy → EKA Literacy	-.058(.069)	.406	-.051(.061)	.404
BKA Numeracy → EKA Numeracy	.362(.077)	<.001	.362(.077)	<.001
BKA Literacy with BKA Numeracy	.505(.070)	<.001	.258(.051)	<.001
EKA Literacy with EKA Numeracy	.327(.082)	<.001	.103(.030)	.001

Note: Dataset underwent multiple imputation (N = 461). Above are the standardized and unstandardized estimates, pooled across 20 datasets. EF refers to Executive Functioning, DR to Developmental Readiness, BKA to Beginning of Kindergarten Assessment, and EKA to End of Kindergarten Assessment.

Model fit indices: $\chi2$ (40) = 110.438, p < .001, comparative fit index = 0.943; Tucker-Lewis index = 0.905; root mean square error of approximation = 0.062; standardized root mean square residual = 0.044).

Control variables: Gender, minority status, SES (geocode), and years from screening until starting kindergarten were assessed as control variables. Only years until starting kindergarten was a significant predictor for EKA numeracy [β = 0.241(.071), p = .001; B = -.310(.091), *p* = .001], so it was retained in the final model. All other control variables were removed.

Moderating variables: Gender, minority status, SES (geocode), and years from screening until starting kindergarten were assessed for having a moderating effect on the predictive strength of EF and DR for outcome variables. There were no significant effects and therefore these paths were not retained in the final model.

Table 9. *Structural Equation Modeling of Executive Function, Developmental Readiness, and Additional Screening Measures as Predictors of Kindergarten Behavior*

Path	Standardized Estimate (Std. Error)	*p*	Unstandardized Estimate (Std. Error)	*p*
DR with EF	.749(.028)	<.001	.422(.047)	<.001
DR with ASQ-SE	-.385(.046)	<.001	-.379(.065)	<.001
DR with MPSI-R Behavior	-.510(.051)	<.001	-1.077(.183)	<.001
EF with ASQ-SE	-.406(.049)	<.001	-.244(.041)	<.001
EF with MPSI-R Behavior	-.380(.057)	<.001	-.490(.090)	<.001
ASQ-SE with MPSI-R Behavior	.307(.071)	<.001	.693(.197)	<.001
DR→ Kindergarten Behavior	.108(.131)	.411	.838(1.039)	.420
EF→ Kindergarten Behavior	-.294(.112)	.009	-3.782(1.479)	.011
ASQ-SE→ Kindergarten Behavior	.186(.079)	.018	1.369(.586)	.019
MPSI-R Behavior → Kindergarten Behavior	.209(.078)	.007	.718(.274)	.009

Note: Dataset underwent multiple imputation (N = 461). Above are the standardized and unstandardized estimates, pooled across 20 datasets. EF refers to Executive Functioning, DR to Developmental Readiness. ASQ-SE, MPSI-R Behavior, Kindergarten Behavior, and First Grade Behavior are scored such that higher scores represent more problem behavior.

Model fit indices: $\chi2(28) = 79.819$; comparative fit index = 0.967; Tucker-Lewis index = 0.948; root mean square error of approximation = 0.061; standardized root mean square residual = 0.031.

Control variables: Gender, minority status, SES (geocode), and years from screening until starting kindergarten were assessed as control variables. There were no significant effects; therefore, these variables were not included in the final model.

Moderating variables: Gender, minority status, SES (geocode), and years from screening until starting kindergarten were assessed for having a moderating effect on the predictive strength of EF, DR, ASQ-SE, and MPSI-R Behavior for outcome variables. There were no significant effects and therefore these paths were not retained in the final model.

Table 10. *Structural Equation Modeling of Executive Function, Developmental Readiness, and Additional Screening Measures as Predictors of Kindergarten and First Grade Behavior*

Path	Standardized Estimate (Std. Error)	*p*	Unstandardized Estimate (Std. Error)	*p*
DR with EF	.749(.028)	<.001	.422(.047)	<.001
DR with ASQ-SE	-.385(.046)	<.001	-.380(.065)	<.001
DR with MPSI-R Behavior	-.510(.051)	<.001	-1.079(.184)	<.001
EF with ASQ-SE	-.406(.049)	<.001	-.244(.040)	<.001
EF with MPSI-R Behavior	-.380(.057)	<.001	-.489(.090)	<.001
ASQ-SE with MPSI-R Behavior				
DR→ Kindergarten Behavior	.109(.131)	.408	.842(1.037)	.417
EF→ Kindergarten Behavior	-.295(.112)	.008	-3.795(1.476)	.010
ASQ-SE→ Kindergarten Behavior	.186(.079)	.018	1.368(.586)	.020
MPSI-R Behavior → Kindergarten Behavior	.209(.078)	.007	.719(.274)	.009
DR→ 1st Grade Behavior	.208(.157)	.185	1.718(1.308)	.189
EF→ 1st Grade Behavior	-.381(.144)	.008	-5.161(2.029)	.011
ASQ-SE→ 1st Grade Behavior	-.004(.108)	.970	-.031(.831)	.971
MPSI-R Behavior → 1st grade Behavior	.219(.122)	.072	.791(.274)	.009

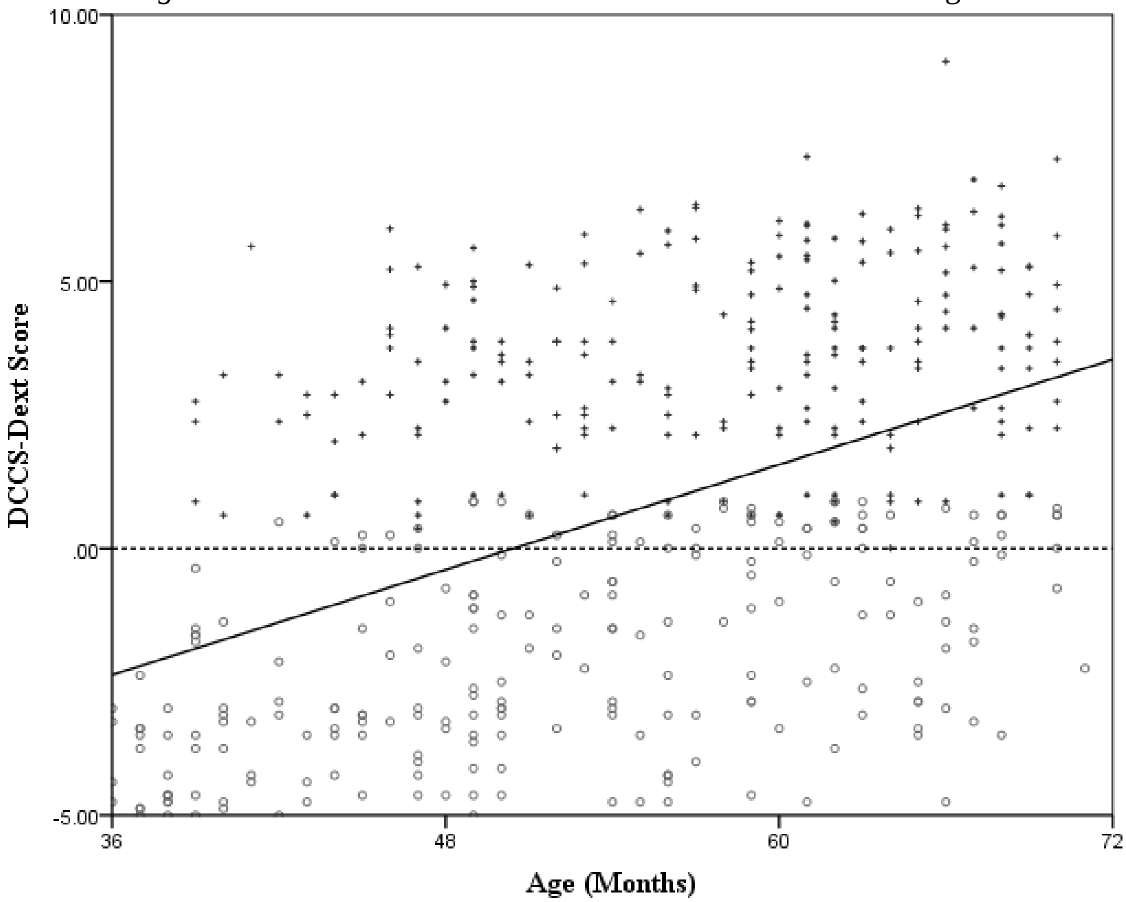
Note: Dataset underwent multiple imputation (N = 461) with the exception of 1st Grade Behavior (N = 165). Above are the standardized and unstandardized estimates, pooled across 20 datasets. EF refers to Executive Functioning, DR to Developmental Readiness. ASQ-SE, MPSI-R Behavior, Kindergarten Behavior, and First Grade Behavior are scored such that higher scores represent more problem behavior.

Model fit indices: $\chi 2$ (33) = 88.994; comparative fit index = 0.996; Tucker-Lewis index = 0.944; root mean square error of approximation = 0.061; standardized root mean square residual = 0.033. Control variables: Gender, minority status, SES (geocode), and years from screening until starting kindergarten were assessed as control variables. There were no significant effects; therefore, these variables were not included in the final model.

Moderating variables: Gender, minority status, SES (geocode), and years from screening until starting kindergarten were assessed for having a moderating effect on the predictive strength of EF, DR, ASQ-SE, and MPSI-R Behavior for outcome variables. There were no significant effects and therefore these paths were not retained in the final model.

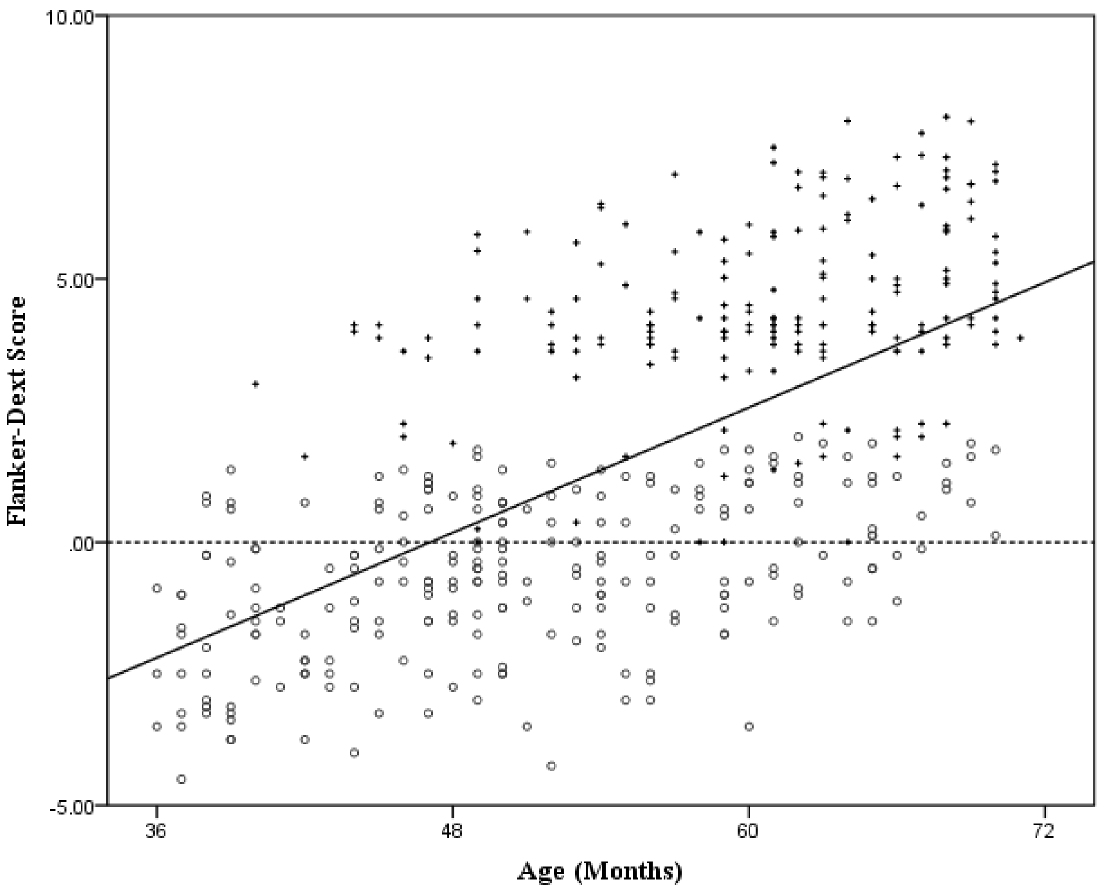
Figures

Figure 1. Scatter-Plot of DCCS-Dext Scores as a Function of Age



Note: Graph demonstrates the variability captured by the developmental extension (dext). All children who scored below 0 required the dext. Solid line reflects correlation of DCCS-Dext with age. All children who scored below 0 required the dext. Some children who score above 0 also are administered the dext; therefore, plus signs were used to present children who were administered the DCCS without dext and circles for children who were administered the DCCS with dext. Sold line reflects correlation of DCCS-Dext with age.

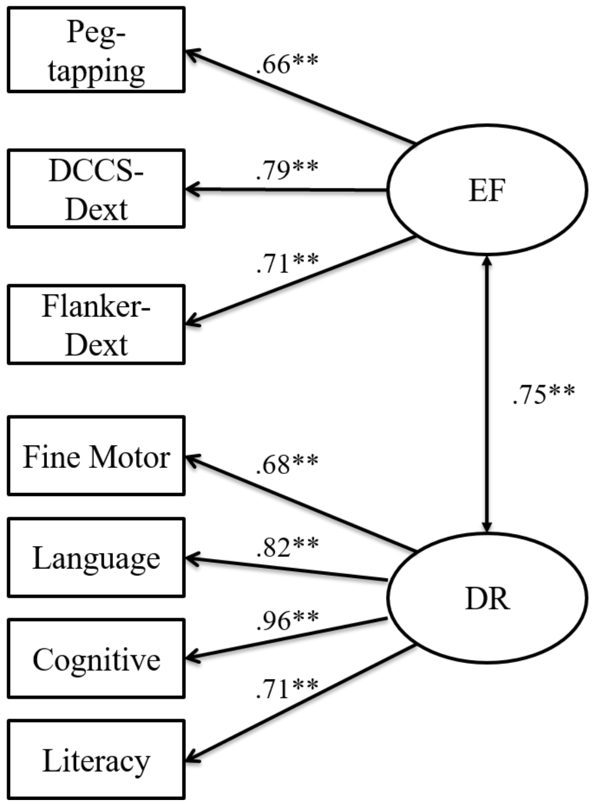
Figure 2. *Scatter-Plot of Flanker-Dext Score as a Function of Age*



Note: Graph demonstrates the variability captured by the developmental extension (dext). All children who scored below 0 required the dext. Some children who score above 0 also are administered the dext; therefore, plus signs were used to present children who were administered the Flanker without dext and circles for children who were administered the DCCS with dext. Sold line reflects correlation of Flanker-Dext with age.

Graph demonstrates a gap in total scores between 2.25 and 3.00. Children above this gap passed TB fish and were therefore given an additional 20 TB arrow items; children below the gap did not perform well enough to be administered this set of items (see Appendix B for more detailed information on scoring).

Figure 3. *Confirmatory Factor Analysis for Executive Function and Developmental Readiness*

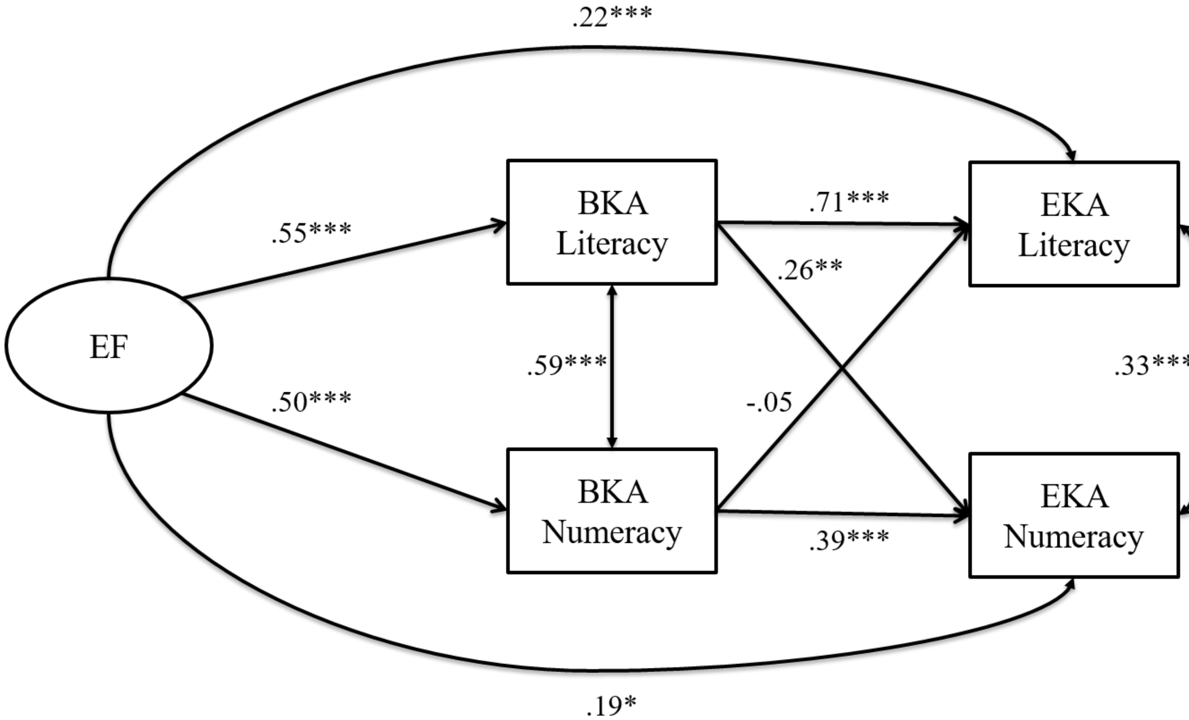


Notes: Indicators are residuals after controlling for age (Peg-Tapping, DCCS-Dext, Flanker-Dext, and literacy controlling for age; fine motor, language, and cognitive controlling for age-squared); EF refers to Executive Function, DR refers to Developmental Readiness

Model fit indices: $\chi2(13) = 32.722$, $p = .002$; comparative fit index = 0.986; Tucker-Lewis index = 0.977; root mean square error of approximation = 0.057; standardized root mean square residual = 0.027.

***p<.001, **p<.01, *p<.05

Figure 4. *Structural Equation Modeling of Executive Function as a Predictor of Kindergarten Literacy and Numeracy*



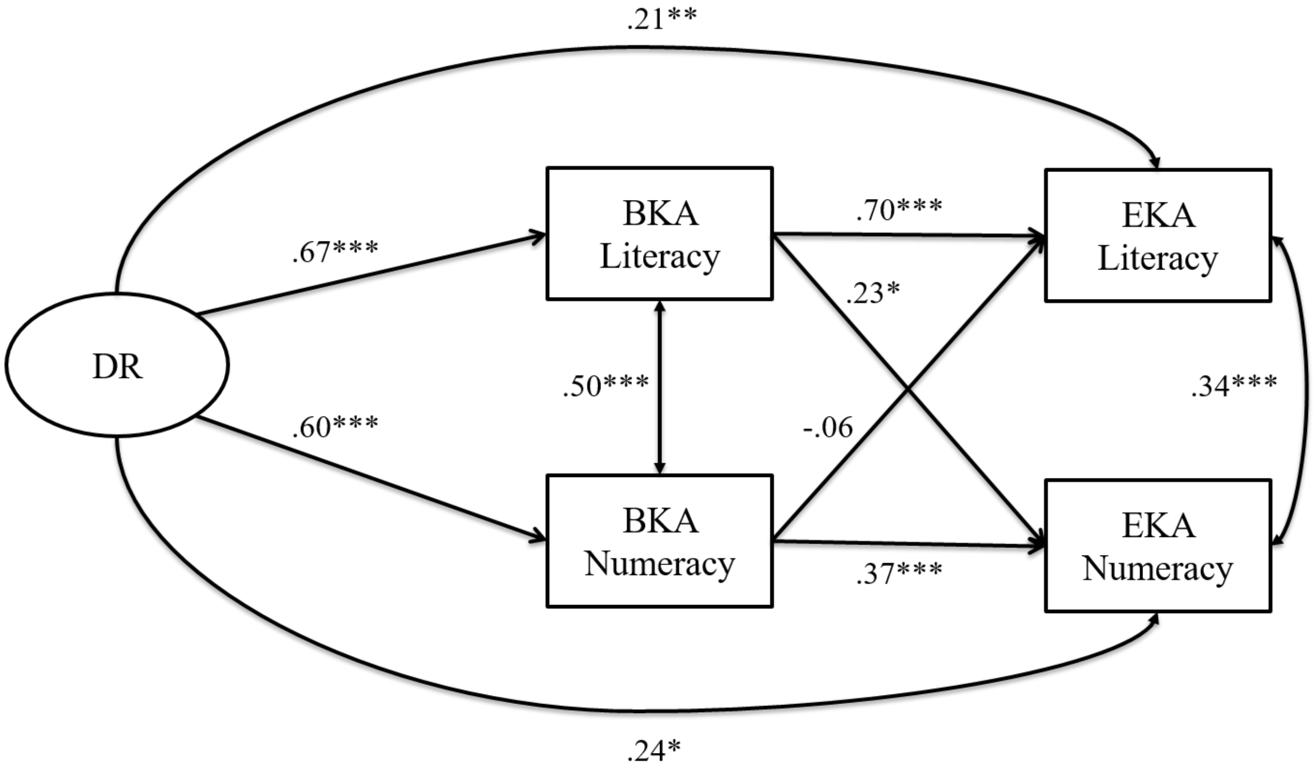
Notes: EF refers to Executive Function; BKA refers to Beginning of Kindergarten Assessment, EKA refers to End of Kindergarten Assessment

Years from time of screening until starting kindergarten was a significant predictor for EKA numeracy [β = -.243(.072), *p* = .001] and was controlled for in the model.

Model fit indices: $\chi 2$ (11) = 14.043, p = .231; comparative fit index = 0.995; Tucker-Lewis index = 0.987; root mean square error of approximation = 0.024; standardized root mean square residual = 0.021.

****p*<.001, ***p*<.01, **p*<.05

Figure 5. *Structural Equation Modeling of Developmental Readiness as a Predictor of Kindergarten Literacy and Numeracy*



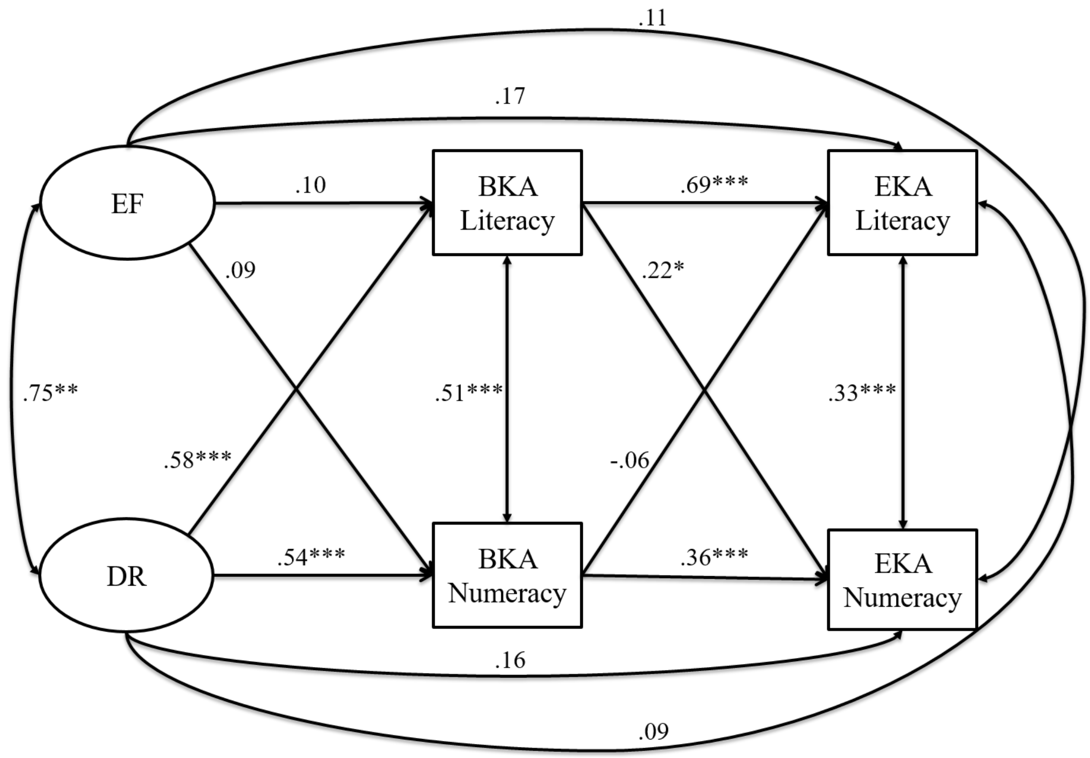
Notes: DR refers to Developmental Readiness, BKA refers to Beginning of Kindergarten Assessment, EKA refers to End of Kindergarten Assessment

Years from time of screening until starting kindergarten was a significant predictor for EKA numeracy [β = -.241(.071), *p* = .001] and was controlled for in the model.

Model fit indices: $\chi 2$ (18) = 59.376, p < .001; comparative fit index = 0.889; Tucker-Lewis index = 0.945; root mean square error of approximation = 0.071; standardized root mean square residual = 0.048.

***p<.001, **p<.01, *p<.05

Figure 6. *Structural Equation Modeling of Executive Function and Developmental Readiness as Predictors of Kindergarten Literacy and Numeracy*



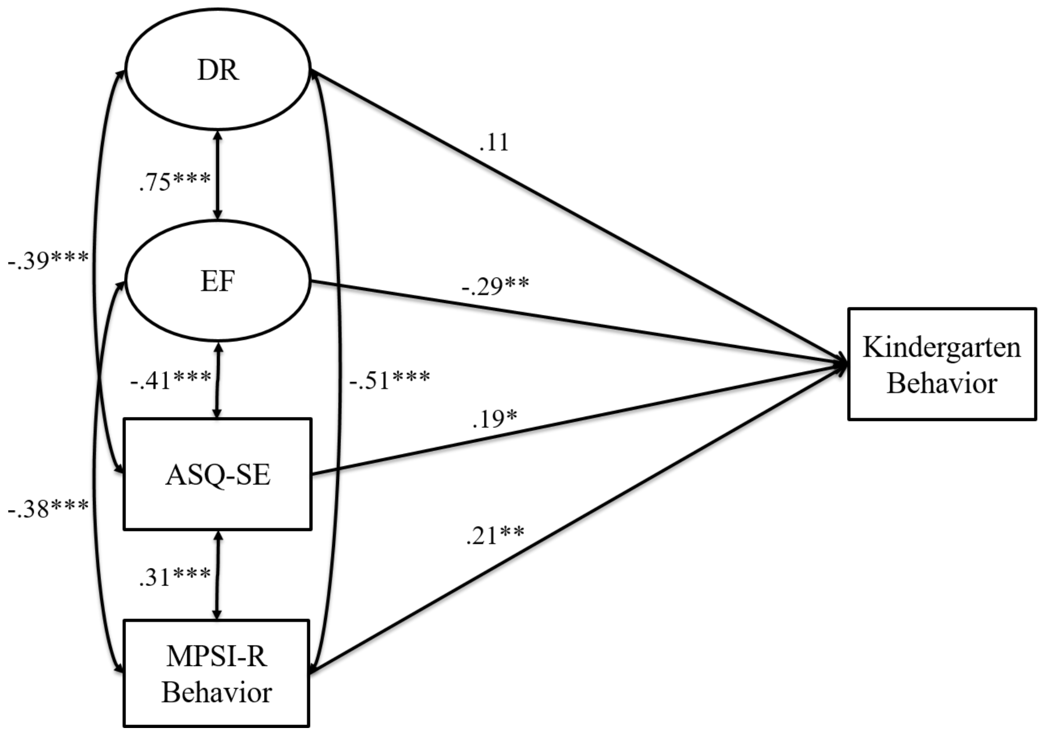
Notes: EF refers to Executive Function, DR refers to Developmental Readiness, BKA refers to Beginning of Kindergarten Assessment, EKA refers to End of Kindergarten Assessment

Years from time of screening until starting kindergarten was a significant predictor for EKA numeracy [β = 0.241(.071), p = .001] and was controlled for in the model.

Model fit indices: $\chi 2$ (40) = 110.438, p < .001, comparative fit index = 0.943; Tucker-Lewis index = 0.905; root mean square error of approximation = 0.062; standardized root mean square residual = 0.044).

****p*<.001, ***p*<.01, **p*<.05

Figure 7. *Structural Equation Modeling of Executive Function, Developmental Readiness, and Additional Screening Measures as Predictors of Kindergarten Behavior*

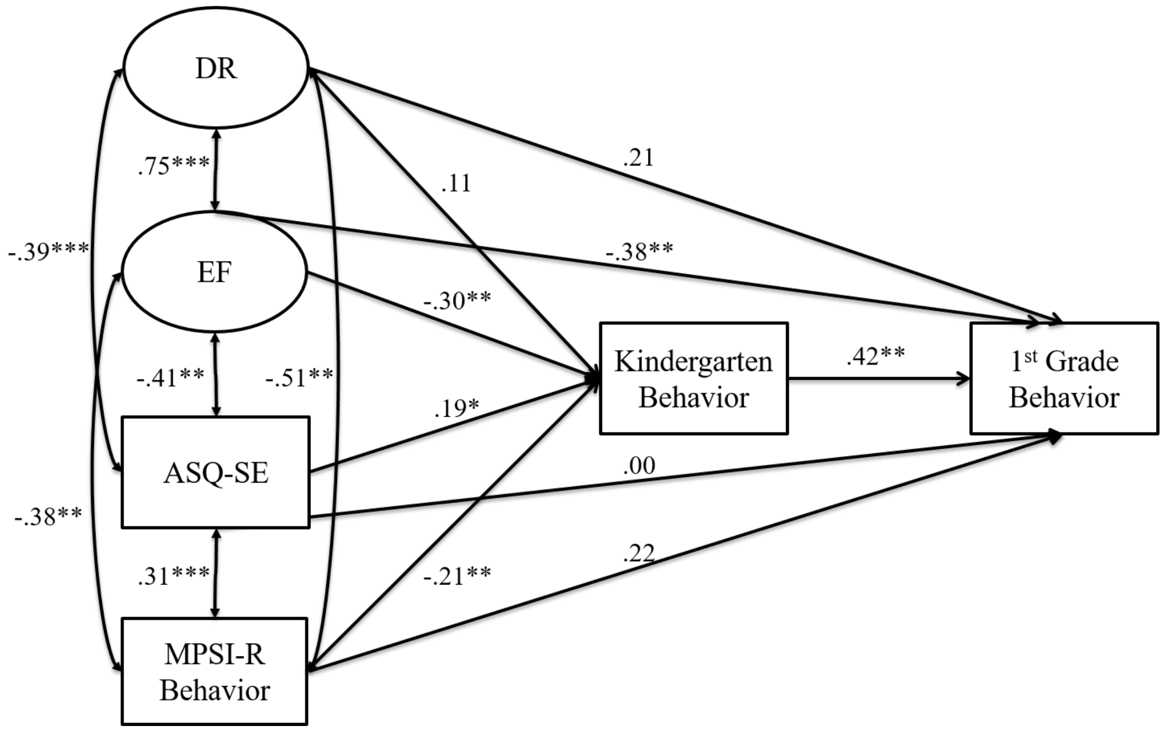


Notes: EF refers to Executive Function, DR refers to Developmental Readiness; ASQ-SE is a parent-report measure of social-emotional concerns; MPSI-R Behavior is measure of screener-rated behavior problems. ASQ-SE, MPSI-R Behavior, and Kindergarten Behavior are scored such that higher scores reflect more problem behavior

Model fit indices: $\chi2(28) = 79.819$; comparative fit index = 0.967; Tucker-Lewis index = 0.948; root mean square error of approximation = 0.061; standardized root mean square residual = 0.031.

****p*<.001, ***p*<.01, **p*<.05

Figure 8. *Structural Equation Modeling of Executive Function, Developmental Readiness, and Additional Screening Measures as Predictors of Kindergarten and First Grade Behavior*



Notes: EF refers to Executive Function, DR refers to Developmental Readiness
ASQ-SE is a parent-report measure of social-emotional concerns; MPSI-Behavior is measure of screener-rated behavior problems. ASQ-SE, MPSI-R Behavior, Kindergarten Behavior, and First Grade Behavior are scored such that higher scores reflect more problem behavior

Model fit indices: $\chi2$ (33) = 88.994; comparative fit index = 0.996; Tucker-Lewis index = 0.944; root mean square error of approximation = 0.061; standardized root mean square residual = 0.033.
****p*<.001, ***p*<.01, **p*<.05

References

Alexander, M. & Stuss, D. (2000). Disorders of frontal lobe functioning, *Seminars in Neurology, 20*, 427–438. doi: 10.1055/s-2000-13175

Anderson, J.E., Zelazo, P.D., Carlson, S.M., Kalstabakken, A.W., & Masten, A.S. (2015). *Technical Report for the Flanker-Developmental Extension*.

Barkley, R.A. (2001). The EF and self-regulation: An evolutionary neuropsychological perspective. *Neuropsychology Review, 11*, 1–29. doi: 10.1023/A:1009085417776

Benson, J., Sabbagh, M., Carlson, S.M., and Zelazo, P.D. (2013). Individual differences in executive functioning predict preschoolers' improvement from theory-of-mind training. *Developmental Psychology, 49*, 1615–1627. doi: 10.1037/a0031056.

Bergman Nutley, S., Soderqvist, S., Bryde, S., Thorell, L.B., Humphreys, K., & Klingberg, T. (2011). Gains in fluid intelligence after training non-verbal reasoning in 4-year-old children: a controlled, randomized study. *Developmental Science, 14*, 591–601. doi: 10.1111/j.1467-7687.2010.01022.x

Bernier, A., Carlson, S. M., & Whipple, N. (2010). From external regulation to self-regulation: Early parenting precursors of young children's executive functioning. *Child Development*, 81, 326-339. doi: 10.1111/j.1467-8624.2009.01397.x

Bierman, K.L., Nix, R.L., Greenberg, M.T., Blair, C., & Domitrovich, C.E. (2008). Executive functions and school readiness interventions: Impact, moderation, and mediation in the Head Start REDI program. *Development and Psychopathology, 20*, 821–843. doi: 10.1017/S0954579408000394

Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American Psychologist, 57*, 111–127. doi: 10.1037/0003-066X.57.2.111

Blair, C & Razza, R.P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78*, 647–680. doi: 10.1111/j.1467-8624.2007.01019.x

Blair, C., & Raver, C.C. (2015). School readiness and self-regulation: A developmental psychobiological approach. *Annual Review of Psychology, 66,* 711–731. doi: 10.1146/annurev-psych-010814-015221

Boan, C. H., Aydlett, L., & Multunas, N. (2007). Early childhood screening and readiness assessment. In R. J. Nagle (Ed.), *Psychoeducational assessment of preschool children* (4th ed., pp. 49–67). Mahwah, NJ: Lawrence Erlbaum

Bodrova, E. & Leong, D.J. (2007). *Tools of the Mind: The Vygotskian Approach to Early Childhood Education, 2nd Edition*. New York: Merrill/Prentice Hall.

Brassard, M. & Boehm, A. (2007). *Preschool assessment: Principles and Practices*. New York: The Guilford Press.

Buckner, J. C., Mezzacappa, E., & Beardslee, W. R. (2003). Characteristics of resilient youths living in poverty: The role of self-regulatory processes. *Development and Psychopathology, 15,* 139–162. doi: 10.1017.S0954579403000087

Bull, R., Espy, K.A., & Wiebe, S.A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology, 33*(3), 205–228. doi: 10.1080/87565640801982312

Bull, R., & Scerif, G. (2001). Executive Functioning as a Predictor of Children's Mathematics Ability: Inhibition, Switching, and Working Memory. *Developmental Neuropsychology*, *19*, 273–293. doi: 10.1207/S15326942DN1903_3

Campbell, F.A., Pungello, E.P., Miller-Johnson, S., Burchinal, M., & Ramey, C.T. (2001). The development of cognitive academic abilities: Growth curves from an early childhood education experiment. *Developmental Psychology*, 37, 231 – 242. doi: 10.1037/0012-1649.37.2.231

Carlson, S.M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology*, *28*, 595–616. doi: 10.1207/s15326942dn2802_3

Carlson, S. M., Zelazo, P. D., & Faja, S. (2013). Executive function. *Oxford Handbook of Developmental Psychology*, *1*, 706–742. doi: 10.1093/oxfordhb/9780199958450.013.0025

Carlson, S.M, Zelazo, P.D., Anderson, J.E., Kalstabakken, A.W. & Masten A.S. (2015). *Technical Report for the Dimensional Change Card Sort- Developmental Extension*.

Carlson, S. M. & Zelazo, P. D. (2014). *Minnesota Executive Function Scale – Test Manual*. Reflection Sciences, LLC. Saint Paul, MN.

Clark, C.A., Pritchard, V.E., & Woodward, L.J. (2010). Preschool executive functioning abilities predict early mathematics achievement. *Developmental Psychology*, 46(5), 1176–1191. doi: 10.1037/a0019672

Cole, P.M., Usher, B.A., & Cargo, A.P. (1993). Cognitive risk and its association with risk for disruptive behavior disorder in preschoolers. *Journal of Clinical Child Psychology*, *22*, 154–164. doi: 10.1207/s15374424jccp2202_3

Cutuli, J. J., Desjardins, C. D., Herbers, J. E., Long, J. D., Heistad, D., Chan, C. K., . . . Masten, A. S. (2013). Academic Achievement Trajectories of Homeless and Highly Mobile Students: Resilience in the Context of Chronic and Acute Risk. *Child Development, 84*(3), 841–857. doi: 10.1111/cdev.12013

Della Sala, S., Gray, C. Spinnler, H., Trivelli, C., (1998). Frontal lob functioning in man: The riddle revisited. *Archives of Clinical Neuropsychology, 12*, 663–682. doi: 10.1016/S0887-6177(97)00093-0

Denckla, M. (1994). Measurement of executive function. In G. Reid Lyon (Ed.), *Frames of reference for the assessment of learning difficulties* (pp. 117 – 142). Blatimore: Brookes Publishing Co.

DePrince, A.P., Weinzierl, K.M., & Combs, M.D. (2009). Executive function performance and trauma exposure in a community sample of children. *Child Abuse Neglect*, *33*, 353–361. doi: 10.1016/j.chiabu.2008.08.002

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. Stuss & R. Knight (Eds.), *Principles of frontal lobe function* (pp. 466–503). New York: Oxford University Press.

Diamond, A. & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, 333, 959–964. doi: 10.1126/science.1204529

Diamond, A., & Ling, D.S. (2016). Conclusions about interventions, programs, and approaches for improving executive functions that appear justified and those that, despite much hype, do not. *Developmental Cognitive Neuroscience*, 18, 34–48. doi: 10.1016/j.dcn.2015.11.005

Diamond, A. & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said to “Do as I say, not as I do.” *Developmental Psychobiology, 29*, 315 – 334. doi: 10.1002/(SICI)1098-2302(199605)29:4<315::AID-DEV2>3.0.CO;2-T

Diamond, A., Prevor, M.B., Callender, G.. & Druin, D.P. (1997). Prefrontal cortex cognitive deficits in children treated early and continuously for PKU. *Monographs of the Sociaty for Research in Child Development, 62*. doi: 10.2307/1166208

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L., . . . Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43,* 1428–1446. doi: 10.1037/0012-1649.43.6.1428

Eisenberg N, Pidada S, & Liew J. (2001). The relations of regulation and negative emotionality to Indonesian children's social functioning. *Child Development, 72*, 1747–1763. doi: 10.1111/1467-8624.00376

Eisenberg, N., Valiente, C., & Eggum, N. (2010). Self-regulation and school readiness. *Early Educational Development, 21*(5), 681–698. doi: 10.1080/10409289.2010.497451

Espinet, S.D., Anderson, J.E., & Zelazo, P.D. (2013). Reflection training improves executive function in preschool children: Behavioral and neural effects. *Developmental Cognitive Neuroscience, 4,* 3–15. doi: 10.1016/j.dcn.2012.11.009.

Espy, K. A., McDiarmid, M.M., Cwik, M. F., Stalets, M. M., Hamby, A., & Senn, T. E. (2004). The contribution of executive functions to emergent mathematic skills in preschool children. *Developmental Neuropsychology, 26,* 465–486. doi: 10.1207/s15326942dn2601_6

Ettenhofer, M. L., Hambrick, D. Z., & Abeles, N. (2006). Reliability and stability of executive functioning in older adults. *Neuropsychology*, 20(5), 607–613. doi: 10.1037/0894-4105.20.5.607

Gredler, G.R. (1997). Issues in early childhood screening and assessment. *Psychology in Schools, 34*(2), 99–106. doi: 10.1002/(SICI)1520-6807(199704)34:2<99::AID-PITS3>3.0.CO;2-N

Hackman, D. A., & Farah, M. J. (2009). Socioeconomic status and the developing brain. *Trends In Cognitive Sciences, 13*(2), 65–73. doi: 10.1016/j.tics.2008.11.003

Hassinger-Das, B., Jordan, N.C., Glutting, J., Irwin, C., & Dyson, N. (2014). Domain-general mediators of the relation between kindergarten number sense and first-grade mathematics achievement. *Journal of Experimental Child Psychology, 118*, 78–92. doi: 10.1016/j.jecp.2013.09.008

Herbers, J. E., Cutuli, J. J., Supkoff, L.M., Heistad, D., Chan, C., Hinz, E., & Masten, A.S. (2012). Early reading skills and academic achievement trajectories of students facing poverty, homelessness, and high residential mobility. *Educational Researcher, 41*(9), 366-374. doi: 10.3102/0013189X12445320

Herbers, J. E., Cutuli, J. J., Supkoff, L. M., Narayan, A. J., &; Masten, A. S. (2014). Parenting and coregulation: Adaptive systems for competence in children experiencing homelessness. *American Journal Of Orthopsychiatry, 84*(4), 420–430. doi: 10.1037/h0099843

Holmes, J., Gathercole, S.E., & Dunning, D.L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science, 12*, F9-F15. doi:10.1111/j.1467-7687.2009.00848.x

Holmes, J., Gathercole, S.E., Place, M., Dunning, D.L., Hilton, K.A., & Elliot, J.G. (2010).Working memory deficits can be overcome: Impacts of training and medication on working memory in children with ADHD. *Applied Cognitive Psychology, 24*, 827–836. doi: 10.1002/acp.1589

Kagan, S. L. (1990). Readiness 2000: Rethinking rhetoric and responsibility. *Phi Delta Kappan*, *72*, 272–279.

Kagan, S.L. (1992). Readiness, past, present, and future: Shaping the agenda. *Young Children, 48*, 48–53.

Kagan, S.L., & Kauerz, K. (2007). Reaching for the whole: Integration and alignment in early education policy. In R. C. Pianta, M. J. Cox, & K. L. Snow (Eds.), *School readiness and the transition to kindergarten in the era of accountability* (pp. 11–30). Baltimore, MD: Paul. H. Brookes.

Kagan, S. and Rigby, E. (February 2003). *Policy Matters: Improving the Readiness of Children for School.* Washington, DC: The Center for the Study of Social Policy.

Kendler, H.H. & Kendler, T.S. (1961). Effect of verbalization on reversal shifts in children. *Science, 134*, 1619–1620. doi: 10.1126/science.134.3490.1619

Klingberg, T., Fernell, E., Olesen, P.J., Johnson, M., Gustafsson, P.,...& Westerberg, H.(2005). Computerized Training of Working Memory in Children With ADHD-A Randomized, Controlled Trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, 44, 177–186. doi: 10.1097/00004583-2005020000-00010

Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences, 14*(7), 317–324. doi: 10.1016/j.tics.2010.05.002

Kochanska, G., Coy, K.C., Murray, K.T. (2001). The development of self-regulation in the first four years of life. *Child Development, 72*(4), 1091–1111. doi: 10.1111/1467-8624.00336

Loman, M.M., Johnson, A.E., Westerlund, A, Pollak, S.D., Nelson, C.A., & Gunnar, M.R. (2012). The effect of early deprivation on executive attention in middle childhood. *Journal of Child Psychology and Psychiatry, 54,* 37–45. doi: 10.1111/j.1469-7610.2012.02602.x

Lonigan, C.J., Burgess, S.R., & Anthony, J.L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent variable longitudinal study. *Developmental Psychology, 36*, 596–613. doi: 10.1037/0012-1649.36.5.596

Masten, A. S., Herbers, J. E., Desjardins, C. D., Cutuli, J. J., McCormick, C. M., Sapienza, J. K., . . . Zelazo, P. D. (2012). Executive Function Skills and School Success in Young Children Experiencing Homelessness. *Educational Researcher, 41*(9), 375–384. doi: 10.3102/0013189x12459883

May, D., & Kundert, D. (1992). Kindergarten screenings in New York State: Tests, purposes, and recommendations. *Psychology in the Schools, 29*, 35–41. doi: 10.1002/1520-6807(199201)29:1<35::AID-PITS2310290108>3.0.CO;2-6

Mazzocco, M.M.M. & Kover, S.T. (2007). A longitudinal assessment of executive function skills and their association with math performance. *Child Neuropsychology, 13*(1), 18–45. doi: 10.1080/09297040600611346

McClelland, M.M., Acock A.C., & Morrison, F.J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly, 21*, 471–490. doi: 10.1016/j.ecresq.2006.09.003

McClelland, M.M., Cameron, C.E., Connor, C.M., Farris, C.L., Jewkes, A.M., & Morrison, F.J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology, 43*, 947–959. doi: 10.1037/0012-1649.43.4.947

McClelland, M. M., Acock, A. C., Piccinin, A., Rhea, S. A., & Stallings, M. C. (2013). Relations between preschool attention span-persistence and age 25 educational outcomes. *Early Childhood Research Quarterly, 28*(2), 314–324. doi: 10.1016/j.ecresq.2012.07.008

Meisels, S.J. (1986). Testing four- and five-year-olds: Response to Salzer and to Shepard and Smith. *Educational Leadership*, *44*(3), 90–92.

Meisels, S.J. (1987). Use and abuses of developmental screening and school readiness testing. *Young Children*, *42,* 68–73.

Meisels, S.J. (1999). Assessing readiness. In R.C. Pianta & M.J. Cox (Eds). *The Transition to Kindergarten* (pp. 39–66). Baltimore: Paul H. Brookes.

Mehaffie, K.E. & McCall, R.B. (2002). *Kindergarten readiness: An overview of the issues and assessment* (Special Report). University of Pittsburgh, Office of Child Development.

Minneapolis Kindergarten Assessments. (December, 2007). *Research, Evaluation, and Assessment; Minneapolis Public Schools. Minneapolis Preschool Screening Instrument – Revised*. (2007). Technical Manual.

Mischel, W., & Patterson, C. J. (1976). Substantive and structural elements of effective plans for self-control. *Journal of Personality and Social Psychology,34*(5), 942. doi: 10.1037/0022-3514.34.5.942

Morrison, F.J., Ponitz, C.C., & McClelland, M.M. (2010). Self-regulation and academic achievement in the transition to school. *Child Development at the Intersection of Emotion and Cognition*, 203–224. doi: 10.1037/12059-011

Muthén, L. K., & Muthén, B. O. (1998–2011). *Mplus User's Guide. Sixth Edition*. Los Angeles, CA: Muthén & Muthén.

National Education Goals Panel (1991). *The national education goals report: building a nation of learners.* Washington, DC: U.S. Government Printing Office.

National Education Goals Panel. (1995). *National Educational Goals report executive summary: Improving education through family-school-community partnerships.* Washington, DC.

National Early Literacy Panel. (2005). *Report on a synthesis of early predictors of reading*. Louisville, KY.

National Research Council. (2008). *Early childhood assessment: Why, what, and how*. Washington, DC: The National Academies Press.

Obradović, J. (2010). Effortful control and adaptive functioning of homeless children: Variable-focused and person-focused analyses. *Journal of Applied Developmental Psychology, 31*(2), 109–117. doi: 10.1016/j.appdev.2009.09.004

Pears, K.C., Fisher, P.A., Bruce, J., Kim, H.K., & Yoerger, K. (2010). Early elementary school adjustment of maltreated children in foster care: The role of inhibitory control and caregiver involvement. *Child Development*, *81*, 1550–1564. doi: 10.1111/j.1467-8624.2010.01491.x

Pechtel, P. & Pizzagalli D.A. (2011). Effects of early life stress on cognitive and affective function: An integrated review of human literature. *Psychopharmacology (Berl), 214*, 55–70. doi: 10.1007/s00213-010-2009-2

Princiotta, D., Flanagan, K. D., and Germino Hausken, E. (2006). *Fifth Grade: Findings From The Fifth Grade Follow-up of the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K)*. (NCES 2006-038) U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Ramey, C.T. & Campbell, F.A. (1991). Poverty, early childhood education, and academic competence: the Abedcedarian experiment. In A. Huston (Ed.), *Children reared in poverty* (pp. 190–221). New York: Cambridge University Press.

Ramey, C.T., Campbell, F.A., Burchinal, M., Skinner, M.L., Gardner, D.M., & Ramey, S.L. (2000). Persistent effects of early intervention on high-risk children and their mothers. *Applied Developmental Science, 4*, 2–14. doi: 10.1207/S1532480XADS0401_1

Raver, C.C., Jones, S.T., Li-Grining, C., Zhai, F., Metzger, M., & Solomon B. (2009). Targeting children's behavior problems in preschool classrooms: A cluster-randomized controlled trial. *Journal of Consulting and Clinical Psychology, 77*, 302–316. doi: 10.1037/a0015302

Rueda, M.R. Fan, J., McCandliss, B.D., Halparin, J.D., Gruber, D.B., Lecari, L.P., & Posner, M.I. (2004). Development of attentional network in childhood. *Neuropsychologia, 42*, 1029 – 1040. doi: 10.1016/j.neuropsychologia.2003.12.012

Rimm-Kaufman, S.E., Pianta, R.C., & Cox, M.J. (2000). Teachers' judgments of problems in the transition to kindergarten. *Early Childhood Research Quarterly, 15*(2), 147–166. doi: 10.1016/S0885-2006(00)00049-1

Rimm-Kaufmann & Pianta, R.C. (2000). An ecological perspective on the transition to kindergarten: A theoretical framework to guide empirical research. *Journal of Applied Developemntal Psychology, 21*(5), 491–511. doi: 10.1016/S0193-3973(00)00051-4

Rydz, D., Shevell, M.I., Majnemer, A., & Oskoui, M. (2005). Topical review: Developmental screening. *Journal of Child Neurology*, *20*(1), 4–21. doi: 10.1177/08830738050200010201

Slotkin, J., Nowinski, C., Hays, R., Beaumont, J., Griffith, J., Magasi, S., Salsman, J. & Gershon, R. (2012). NIH Toolbox Scoring and Interpretation Guide.

Social Explorer Professional, Social Explorer. Accessed July 30, 2014: https://www.lib.umn.edu/indexes/moreinfo?id=14592

Squires, J., Bricker, D., & Twombly, E. (with Yockelson, S., Davis, M.S., Kim, Y.). (2002). *Ages & Stages Questionnaire: Social-Emotional (ASQ:SE): A parent-completed, child-monitoring system for social-emotional behaviors.* Baltimore: Paul H. Brookes Publishing Co.

St. Clair-Thompson, H.L., & Gathercole, S.E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology*, *59*(4), 745–759. doi: 10.1080/17470210500162854

Thorell, L.B., Lindqvist, S., Bergman Nutley, S., Bohlin, G., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science*, 12, 106–113. doi: 10.1111/j.1467-7687.2008.00745.x

Vitaro, F., Brendgen, M., Larose, S., & Tremblay, R.E. (2005). Kindergartendisruptive behaviors, protective factors, and educational achievement by early adulthood. *Journal of Educational Psychology, 97*, 617–629. doi: 10.1037/0022-0663.97.4.617

Willoughby, M. & Blair, C. (2011). Test-retest reliability of a new executive function battery for use in early childhood. *Child Neuropsychology, 17*(6), 564–579. doi: 10.1080/09297049.2011.554390

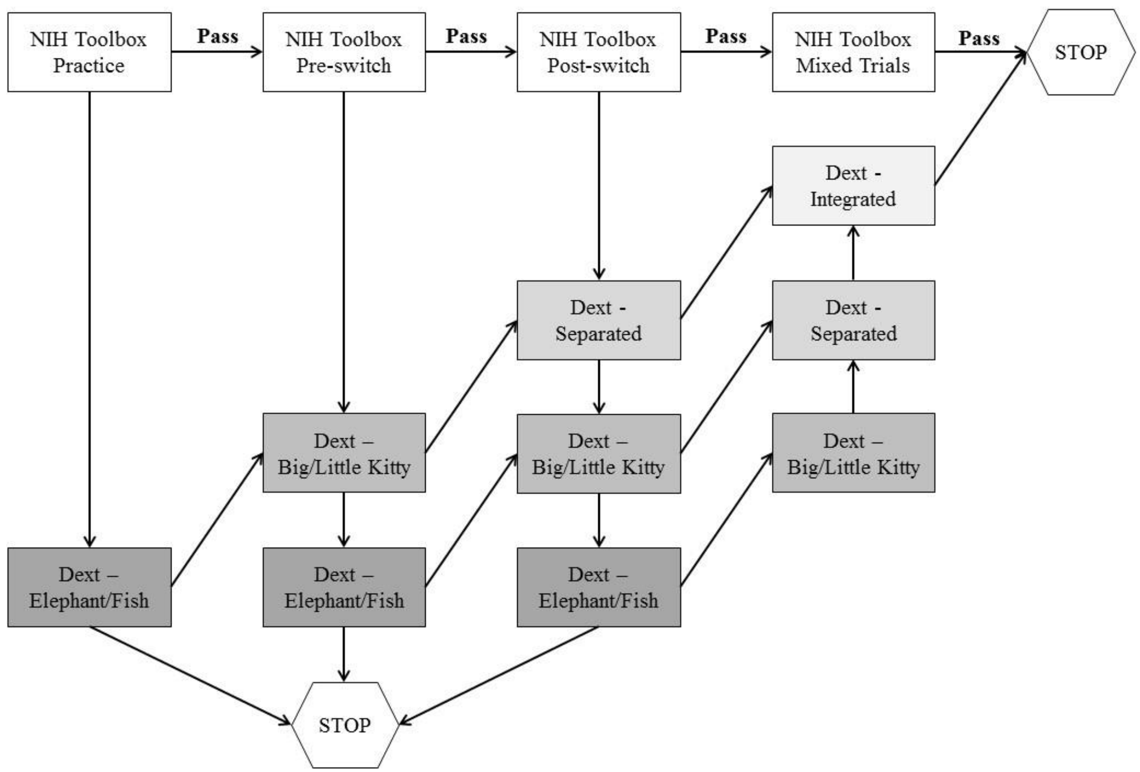
Wilson, B.J. (2003). The role of attentional processes in children's prosocial behavior with peers: Attention shifting and emotion. *Development and Psychopathology*, 15, 313–329. doi: 10.1017/S0954579403000178

Zaitchik, D., Iqbal, Y., & Carey, S. (2014). The effect of executive function on biological reasoning in young children: An individual differences study. *Child Development*, 85(1), 160–75. doi: 10.1111/cdev.12145

Zelazo, P.D. (2006). The Dimensional Change Card Sort (DCCS): a method of assessing executive function in children. *Nature Protocols, 1*, 297 – 301. doi: 10.1038/nprot.2006.46

Zelazo, P. D., & Muller, U. (2002). *Executive function in typical and atypical development*. In U. Goswami (Ed.), Handbook of childhood cognitive d development (pp. 445 – 469). Oxford, UK: Blackwell.

Appendix A: Dimensional Change Card Sort with Developmental Extension Flow Chart



Note: Possible paths between the NIH Toolbox version and the developmental extension (dext) are demonstrated. Note that in the developmental extension, easier modules are administered until the child reaches a baseline (i.e., lowest level child can pass). Upward arrows represent next level if level passed; downward arrows represent next level if child fails. No level is ever administered more than once.

Description of Dimensional Change Card Sort with Developmental Extension

The DCCS – Dext is a computerized, touchscreen measure of cognitive flexibility. On this task, children are presented with bivalent test pictures (e.g., white rabbit) in the middle of the screen and two bivalent target pictures (e.g., white boat, brown rabbit) towards the bottom. Children are given instructions on the dimension by which they should sort the test pictures. This measure is a modified version of the NIH Toolbox DCCS (see Slotkin et al., 2012 for the NIH Toolbox Scoring and Interpretation Guide).

On this task, all children begin with the NIH Toolbox practice trials. Practice begins by introducing children to the stimuli and the first sorting dimension (i.e., shape). Children are then administered four trials and are provided computer-automated feedback on each trial regarding whether responses are correct or incorrect. If children fail (i.e., get less than three trials correct), the sorting rules for the first dimension are provided again and the children are given an additional set of four trials. Children who fail a second time move to the developmental extension (Level 1: Elephant/Fish). If children pass either the first or second attempt, they begin the next part of practice. Children are introduced to the second sorting dimension (i.e., color) and are administered four practice trials for this sorting rule. Pass/fail criteria are the same as for the first dimension. Children continue to receive computer-automated feedback. If they fail, children are reminded of the rules and are given an additional set of four practice trials. Children who do not meet pass criteria on the second attempt move to the developmental extension (Level 1: Elephant/Fish), and children who pass move onto the NIH Toolbox pre-switch items.

On the NIH Toolbox pre-switch trials, children are provided with five trials and are asked to sort by the dimension with which they ended the practice trials (i.e., color). If children are correct on four or more trials, they are administered the NIH Toolbox post-switch trials. If children get fewer than four trials correct, they move to the developmental extension (Level 2: Big/Little Kitty).

On the NIH Toolbox post-switch trials, the sorting is switched (i.e., shape), and children are asked to sort the stimuli using the new rule. Children pass if they get four or more items correct, and fail if they get fewer. Children who pass go onto the NIH Toolbox mixed trials, which consist of 30 trials that involve sorting by both color and shape. After children finish the mixed trials, the task is complete. Children who fail the NIH Toolbox post-switch trials move to the developmental extension (Level 3: Separated DCCS).

Developmental Extension Levels

The developmental extension, which is based on Carlson and Zelazo (2014), consists of four levels: Elephant/Fish (Level 1), Big/Little Kitty (Level 2), Separated DCCS (Level 3), and Integrated DCCS (Level 4). The lower levels are easier and involve sorting simpler objects with less cognitive interference. Additional information on each level is provided below:

Elephant/Fish: Children are presented with elephant and fish test pictures and target pictures. Pre-switch: children are instructed to sort by selecting the elephant target when they see an elephant test card. Post-switch: children are instructed to sort by selecting the fish target card when they see a fish test card.

Big/Little Kitty: Children are presented with “big kitty” and “little kitty” test pictures and target pictures. Pre-switch: children are instructed to sort by selecting the big kitty target card when they see the big kitty test card and the little kitty target card when they see the little kitty test card. Post-switch: children are instructed to “play a silly game” and to sort by selecting the big kitty target card when they see the little kitty test card and the little kitty target card when they see the big kitty test card

Separated DCCS: Children are presented with bivalent (shape, color) test and target cards. This version is similar to the NIH Toolbox DCCS, except that the color is outside of the shape, which provides less cognitive interference and is therefore easier. Pre-switch: children are instructed to sort the test cards by selecting the target card that matches the same color. Post-switch: children are instructed to sort the test cards by selecting the target card that matches the same shape.

Integrated DCCS: Children are presented with bivalent (shape, color) test and target cards. This version uses the same type of target and test cards as the NIH Toolbox DCCS. Pre-switch: children are instructed to sort the test cards by selecting the target card that matches the same color. Post-switch: children are instructed to sort the test cards by selecting the target card that matches the same shape.

Moving between DCCS – Levels

Each level consists of five pre-switch and five post-switch items. Children who pass the pre-switch condition (i.e., four or more trials correct) are administered the post-switch.

Children who fail either pre-switch or post-switch move to the next lower level in the developmental extension, while children who pass both move up to the next level.

The flow chart at the beginning of Appendix A demonstrates how children can move between levels. Note that children continue to receive lower levels of the developmental extension until they have reached their baseline (i.e., the highest level for which they meet pass criteria for both pre-switch and post-switch). They continue to advance in the task until they can no longer pass higher levels. Children are never administered a level more than once.

Scoring the Developmental Change Card Sort with Developmental Extension

Possible scores on the DCCS range from -5.00 to 10.00. All children who have scores below 0.00 required the developmental extension; however, it is possible to have a positive score and still have received the developmental extension.

Possible scores based on accuracy alone are as follows:

Accuracy = -5 + .125*(Correct Dext Items) + .125*(#Correct TB Items)

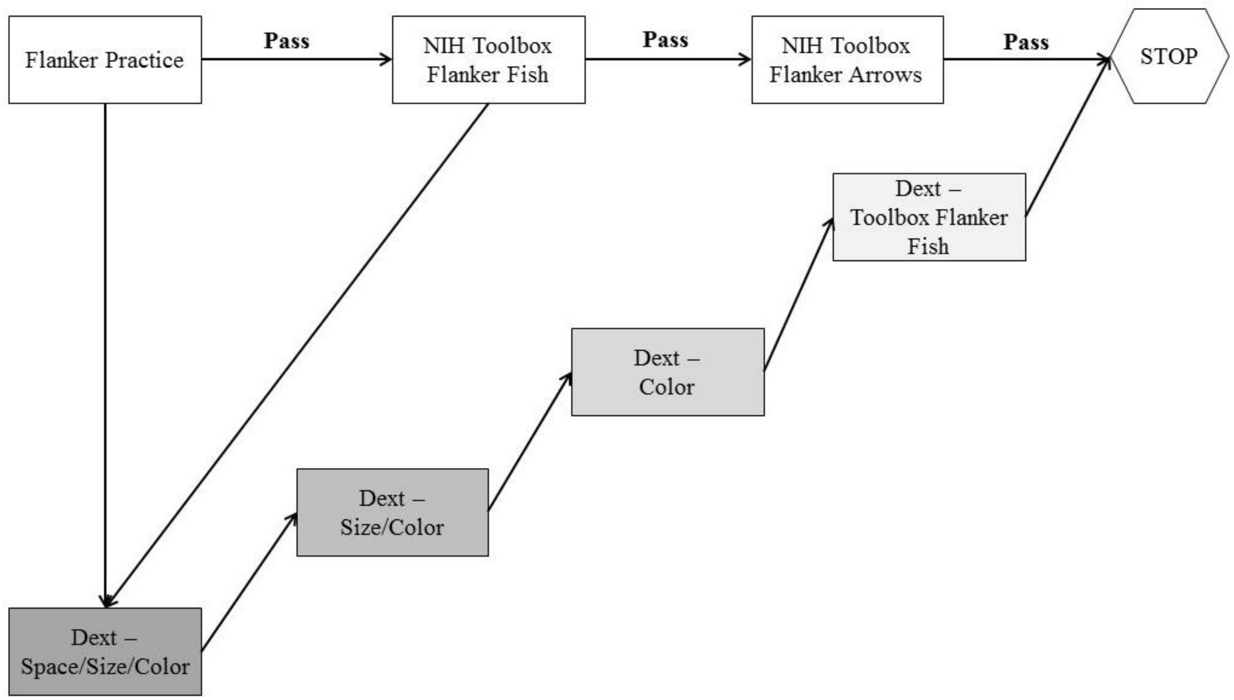
Children who earned an accuracy score greater than 80% may earn additional points for reaction time. The algorithm for calculating reaction time score is as follows:

Reaction Time Score = 5 – (5* [(logRT – log(500)) / (log(3000) – log(500))])

For more information on calculating reaction time score see the NIH Toolbox Scoring and Interpretation Guide (Slotkin et al., 2012).

The total score for the task is the sum of the accuracy and reaction time scores.

Appendix B: Flanker Developmental Extension Flow Chart



Note: Possible paths between the NIH Toolbox version and the developmental extension (dext) are demonstrated. Note that in the developmental extension, all children have the same starting point and continue to advance regardless of performance.

Description of Flanker with Developmental Extension

The Flanker – Dext is a computerized, touchscreen measure of attention and inhibitory control. On this task, children are presented with fish stimuli and are told to focus on the direction that the middle fish is pointing while ignoring the distractor, flanker fish. Children are told that the middle fish is hungry and in order to feed it, they need to select the response button that matches the way that the middle fish is pointing. Sometimes the middle fish points the same direction as the flanker fish (congruent) and sometimes it is pointing in the opposite direction (incongruent). This measure is a modified version of the NIH Toolbox DCCS (see Slotkin et al., 2012 for NIH Toolbox Scoring and Interpretation Guide).

All children begin with the NIH Toolbox Practice trials. Practice begins by introducing children to the stimuli and the rules for the task. Next children are administered four practice trials consisting of two congruent and two incongruent trials. Children are provided computer-automated feedback on whether they are correct or incorrect. If children fail (less than three correct), the instructions are presented again and they are given four new practice trials. If children pass (three or more correct) on either attempt they move onto the NIH Toolbox Flanker Fish. Children who are unable to pass on either set of practice trials move to the developmental extension.

On the NIH Toolbox Flanker Fish, children are administered 20 trials (13 congruent, 7 incongruent). Children pass if they are correct on six or more of the incongruent trials. Those children then move onto NIH Toolbox Flanker Arrows, in which they do a similar task but more challenging task with arrows (i.e., selecting the direction that the middle arrow is pointing, while ignoring flanker arrows). Children who fail the NIH Toolbox Flanker Fish move to the developmental extension.

Developmental Extension Levels

As a part of the developmental extension, children are provided additional instructions, demonstrations, and practice. This component includes “middle practice,” in which the children explicitly practice identifying the middle fish. The child target response buttons are altered so they are an exact match to the arrow inside the fish test stimuli. Further, for all items in this module, a noise that mimics the sound of a fish eating plays after the child selects a target response (regardless of whether the child is correct or incorrect). This was added in order to increase child motivation and interest in the task.

The developmental extension consists of four levels, each with five trials (2 congruent, 3 incongruent). In Level 1, which is intended to be the easiest level, the middle fish is larger than the flanker fish. Additionally, the flanker fish have increased spacing and are a different color from the middle fish. In Level 2, the middle fish continues to be larger and a different color than the flankers; but the original spacing returns. In Level 3, the middle and flanker fish have the original size and spacing but continue to be different colors. Lastly, the stimuli in Level 4 are the same as the original flanker fish.

Moving between Dext Levels

All children start the developmental extension at Level 1 and progress through each level in sequential order, regardless of performance, until the last level has been completed. The flow chart at the beginning of Appendix B demonstrates the possible pathways.

Scoring the Flanker with Developmental Extension

Possible scores on the Flanker with Dext range from -5.00 to 10.00. All children who have scores below 0.00 required the developmental extension; however, it is possible to have a positive score and still have received the developmental extension.

Possible scores based on accuracy alone are as follows:

Accuracy = -5 + .250*(Correct Dext Items) + .125*(#Correct TB Items)

Children who earned an accuracy score greater than 80% may earn additional points for reaction time. The algorithm for calculating reaction time score is as follows:

Reaction Time Score = 5 – (5* [(logRT – log(500)) / (log(3000) – log(500))])

For more information on calculating reaction time score see the NIH Toolbox Scoring and Interpretation Guide (Slotkin et al., 2012).

The total score for the task is the sum of the accuracy and reaction time scores